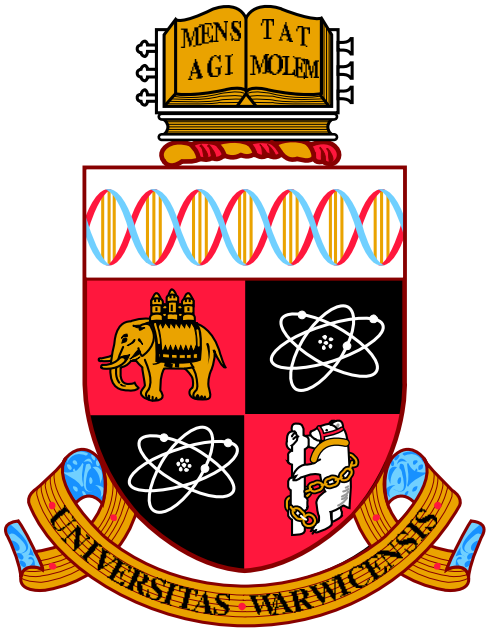



# Point process modelling of coordinate-based meta-analysis neuroimaging data

by

**Pantelis Samartsidis**

**Thesis**

Submitted to the University of Warwick

for the degree of

**Doctor of Philosophy**

**Department of Statistics**

July 2016

THE UNIVERSITY OF WARWICK

***Στα κορίτσια (Μαρία, Τατιάνα, Ειρήνη)***

*-... Χάρηκα πάρα πολύ κύριε Βλαμμένε!*
*-Μη με λέτε βλαμμένο.*
*-Αφού Βλαμμένο σας λένε, πως να σας πώ;*
*-Δόκτωρ.*
*-Αααα, δόκτωρ. Καθήστε δόκτωρ! ...*

---

Διάλογος από '**Της Ελλάδος τα παιδιά**'

# Contents

**List of Tables** **iv**

**List of Figures** **v**

**Acknowledgments** **viii**

**Declarations** **ix**

**Abstract** **x**

**Chapter 1 Introduction** **1**

**Chapter 2 Background** **4**

2.1 Neuroimaging background . . . . . 4

2.1.1 Limitations of individual studies and meta-analysis . . . . . 7

2.2 Spatial point processes background . . . . . 8

**Chapter 3 The coordinate-based meta-analysis of fMRI data: a review** **11**

3.1 Introduction . . . . . 11

3.2 CBMA methods . . . . . 12

3.2.1 Kernel-based methods . . . . . 12

3.2.2 Model-based Methods . . . . . 16

3.3 Evaluation of existing methods . . . . . 21

3.3.1 ALE simulation study . . . . . 21

3.3.2 Analysis of a real dataset . . . . . 25

3.3.3 Discussion . . . . . 28

3.4 Open problems . . . . . 30

**Chapter 4 A Bayesian log-Gaussian Cox process model for CBMA meta-regression** **32**
4.1 Introduction . . . 32
4.2 Model specifications . . . 33
4.2.1 Posterior approximation . . . 35
4.3 Sampling algorithm details . . . 36
4.4 Simulation studies . . . 39
4.4.1 Setup 1 . . . 39
4.4.2 Setup 2 . . . 40
4.5 Application: meta-analysis of emotion and executive control studies 44
4.5.1 Data despcription . . . 44
4.5.2 Algorithm details and convergence diagnostics . . . 44
4.5.3 Results . . . 47
4.5.4 Model assessment . . . 52
4.6 Discussion . . . 52

**Chapter 5 A Bayesian spatial model for group fMRI studies** **55**
5.1 Introduction . . . 55
5.2 The model . . . 56
5.3 Posterior inferences . . . 61
5.4 Simulation studies . . . 65
5.4.1 Simulation setup . . . 65
5.4.2 Analysis of a single dataset . . . 66
5.4.3 Sensitivity to $m$ . . . 73
5.5 Application . . . 74
5.5.1 Data description . . . 74
5.5.2 Implementation details . . . 78
5.5.3 Results . . . 80
5.6 Discussion . . . 84

**Chapter 6 Estimating the number of missing studies in neuroimaging meta-analysis** **86**
6.1 Introduction . . . 86
6.2 The BrainMap database . . . 88
6.3 A zero-truncated count model for CBMA file drawer . . . 89
6.4 Effect of missing studies . . . 94
6.5 Discussion . . . 95

**Chapter 7 Conclusions** **97**
7.1 Contributions . . . . . 97
7.2 Future work . . . . . 98

**Appendix A LGCP supplements** **100**
A.1 Gradient expressions for the LGCP . . . . . 100
A.2 LGCP simulation setup I traceplots . . . . . 104
A.3 Real data analysis diagnostics . . . . . 106
A.4 Full brain analysis . . . . . 112
A.5 Model assessment . . . . . 119

**Appendix B Group fMRI supplements** **122**
B.1 Posterior distribution . . . . . 122
B.2 Sampling algorithm details . . . . . 123
B.3 Real data analysis supplementary figures . . . . . 134

**Appendix C File drawer supplements** **148**
C.1 Zero-truncated regression supplements . . . . . 148
C.2 Emotion CBMA with missing studies . . . . . 151
C.3 Zero-truncated Negative Binomial simulation study . . . . . 153

# List of Tables

3.1 Extract from a CBMA dataset of emotion studies . . . . . 12

4.1 LGCP simulation study 1 results: posterior summaries . . . . . 40

5.1 Group fMRI simulation study: setup for study centers . . . . . 66

6.1 Estimated posterior prevalence of zero-count studies for subsamples A-E . . . . . 91

A.1 Meta-analysis of emotion and executive control studies: ROI analysis for emotions . . . . . 113

A.2 Meta-analysis of emotion and executive control studies: ROI analysis for executive control . . . . . 116

C.1 Zero-truncated Negative Binomial simulation study: bias of $\hat{p}_0$ . . . 153

C.2 Zero-truncated Negative Binomial simulation study: variance of $\hat{p}_0$ . 154

# List of Figures


2.1 An average brain in MNI space . . . . . 5

3.1 Graphical representation of the BHICP . . . . . 19
3.2 ALE simulation study: power against the proportion of valid studies 23
3.3 ALE simulation study: power against the total number of valid studies 24
3.4 Results of the meta-analysis of 164 emotion studies using ALE, MKDA, SDM and the BHICP . . . . . 27

4.1 LGCP simulation study 1: type 1 true and estimated latent Gaussian fields in slices $z = -22$ and $z = 4$ . . . . . 41
4.2 LGCP simulation study 1: type 2 true and estimated latent Gaussian fields in slices $z = -22$ and $z = 4$ . . . . . 42
4.3 LGCP simulation study 2: estimated latent Gaussian fields and mean standardised posterior difference in slice $z = -24$ . . . . . 45
4.4 LGCP simulation study 2: estimated latent Gaussian fields and mean standardised posterior difference in slice $z = -16$ . . . . . 46
4.5 Graphical representation of the meta-analysis dataset consisting of 855 emotion and 338 executive control studies . . . . . 47
4.6 Meta-analysis of emotion and executive control studies: median posterior intensity for emotions . . . . . 49
4.7 Meta-analysis of emotion and executive control studies: median posterior intensity for executive control . . . . . 50
4.8 Meta-analysis of emotion and executive control studies: mean standardised posterior difference of the two types . . . . . 51
4.9 Meta-analysis of emotion and executive control studies: estimated posterior probabilities of observing at least one focus for several ROIs 53

5.1 Realisation of the group fMRI model for 3 participants . . . . . 62
5.2 Group fMRI simulation study: participant 1 results . . . . . 68

5.3 Group fMRI simulation study: participant 8 results . . . . . 69
5.4 Group fMRI simulation study: traceplots for $n\left(\mathbf{x}_1^+\right)$, $n\left(\mathbf{x}_1^-\right)$, $n\left(\mathbf{x}_8^+\right)$, $n\left(\mathbf{x}_8^-\right)$, $n\left(\mathbf{z}^+\right)$ and $n\left(\mathbf{z}^-\right)$ . . . . . 71
5.5 Group fMRI simulation study: voxel-wise posterior probability of observing a study center . . . . . 72
5.6 Group fMRI simulation study: component sensitivity to $m$ . . . . . 75
5.7 Group fMRI simulation study: center sensitivity to $m$ . . . . . 76
5.8 Group fMRI simulation study: true positive rate sensitivity to $m$ . . 77
5.9 Group fMRI study on faces task: center sensitivity analysis . . . . . 81
5.10 Group fMRI study on faces task: voxel-wise posterior probability of observing a study center . . . . . 82
5.11 Group fMRI study on faces task: participant 1 results . . . . . 83

6.1 Empirical distribution of the total number of foci per experiment in the BrainMap database . . . . . 88
6.2 Zero-truncated Poisson and Negative Binomial fit for subsample A . 92
6.3 Estimated posterior prevalence of zero-count studies as function of study year and sample size for subsample A . . . . . 92
6.4 Estimated prevalence of zero-count studies as a function of study context for subsample A . . . . . 93
6.5 Estimated posterior probabilities of observing at least one activation the right and left amygdala, before and after the inclusion of zero-count studies . . . . . 95

A.1 LGCP simulation study 1: traceplots for $\sigma_k$ and $\rho_k$ . . . . . 104
A.2 LGCP simulation study 1: traceplots for $\mu_k$ and $\beta_k$ . . . . . 105
A.3 Meta-analysis of emotion and executive control studies: traceplots and ACFs for $\sigma_k$ . . . . . 106
A.4 Meta-analysis of emotion and executive control studies: traceplots and ACFs for $\rho_k$ . . . . . 107
A.5 Meta-analysis of emotion and executive control studies: traceplots and ACFs for $\mu_k$ . . . . . 108
A.6 Meta-analysis of emotion and executive control studies: traceplots and ACFs for $\int_{\mathcal{B}} \lambda^k(\xi)\, d\xi$ . . . . . 109
A.7 Meta-analysis of emotion and executive control studies: traceplots and ACFs for $\lambda^k_{v_1}$ . . . . . 110
A.8 Meta-analysis of emotion and executive control studies: traceplots and ACFs for $\lambda^k_{v_2}$ . . . . . 111

A.9 Meta-analysis of emotion and executive control studies: posterior predictive check on first order properties . . . . . 120
A.10 Meta-analysis of emotion and executive control studies: posterior predictive check on second order properties . . . . . 121

B.1 Group fMRI study on faces task: traceplots for $n\left(\mathbf{x}_i^+\right)$ and $n\left(\mathbf{x}_i^-\right)$, $i = 1, \ldots, 3$ . . . . . 136
B.2 Group fMRI study on faces task: traceplots for $n\left(\mathbf{x}_i^+\right)$ and $n\left(\mathbf{x}_i^-\right)$, $i = 4, \ldots, 6$ . . . . . 137
B.3 Group fMRI study on faces task: traceplots for $n\left(\mathbf{z}^+\right)$, $n\left(\mathbf{z}^-\right)$, $s_0^2$, $m_0$, $\beta_{s_+^2}$, $m_0$, and $\beta_{s_-^2}$ . . . . . 138
B.4 Group fMRI study on faces task: traceplots for $\mu^+$, $\sigma^2_{m^+}$, $\mu^-$ and $\sigma^2_{m^-}$ 139
B.5 Group fMRI study on faces task: posterior voxel-wise probability of observing an increase center with a different $m$ . . . . . 140
B.6 Group fMRI study on faces task: posterior voxel-wise probability of observing an increase center with a different $\eta$ . . . . . 141
B.7 Group fMRI study on faces task: total number of participants with missing data per voxel . . . . . 142
B.8 Group fMRI study on faces task: participant 2 results . . . . . 143
B.9 Group fMRI study on faces task: participant 3 results . . . . . 144
B.10 Group fMRI study on faces task: participant 4 results . . . . . 145
B.11 Group fMRI study on faces task: participant 5 results . . . . . 146
B.12 Group fMRI study on faces task: participant 6 results . . . . . 147

C.1 Zero-truncated Poisson and Negative Binomial fit for subsamples B-E 149
C.2 Estimated posterior prevalence of zero-count studies as a function of study year and sample size for subsamples B-E . . . . . 150
C.3 Estimated posterior prevalence of zero-count studies as a function of study context for subsamples B-E . . . . . 151
C.4 Meta-analysis of 855 emotion studies adding 86 zero-count studies: median posterior intensity . . . . . 152

# Acknowledgments

First and foremost, I would like to express my deep gratitude towards my family for their unconditional love, endless support and constant encouragement during these past four years. Without them I wouldn't have made it this far.

I would like to thank my supervisor Thomas Nichols for his tremendous knowledge, guidance, motivation and patience. Our meetings have helped me improve not only as a statistician, but as a person as well. I also wish to thank Timothy Johnson for his valuable inputs on my work and the numerous interesting discussions about statistics. Special thanks are due to Athanassios Yannacopoulos, Vassilis Vasdekis, Petros Dellaportas and Takis Besbeas for their wonderful lectures, their help during my undergraduate studies and for encouraging me to continue for a PhD.

I am grateful to my flatmates Panayiota, Apostolis and Thodoris as well as our neighbours Giorgos, Kyriaki and Loukia for all the fun, understanding, patience and help that they have generously given. They are like a family to me. Many thanks to my friends Habib, Simone, Giacomo, Silvia, Ioanna, Stratos, Angelos and Fiona for all the helpful discussions regarding our research and the great times that we shared together. Special thanks go to Helen and Murray for proofreading my thesis.

I would like to thank all members of the Department of Statistics for creating such a friendly and stimulating environment during my studies here at Warwick. Finally, acknowledgements are due to NIH for funding my PhD.

# Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself except where stated and has not been submitted in any previous application for any degree.

- The data for Chapter 4 were provided by Tor Wager and Lisa Feldman Barrett
- The data for Chapter 6 were obtained from the BrainMap database with help from Angela Laird and Peter Fox
- The works in Chapters 3 and 4 are completed and will be submitted for publication soon

# Abstract

Now over 25 years old, functional magnetic resonance imaging (fMRI) has made significant contributions in improving our understanding of the human brain function. However, some limitations of fMRI studies, including those associated with the small sample sizes that are typically employed, raise concerns about validity of the technique. Lately, growing interest has been observed in combining the results of multiple fMRI studies in a meta-analysis. This can potentially address the limitations of single experiments and raise opportunities for reaching safer conclusions. Coordinate-based meta-analyses (CBMA) use the peak activation locations from multiple studies to find areas of consistent activations across experiments. CBMA presents statisticians with many interesting challenges. Several issues have been solved but there are also many open problems. In this thesis, we review literature on the topic and after describing the unsolved problems we then attempt to address some of the most important. The first problem that we approach is the incorporation of study-specific characteristics in the meta-analysis model known as meta-regression. We propose an novel meta-regression model based on log-Gaussian Cox processes and develop a parameter estimation algorithm using the Hamiltonian Monte Carlo method. The second problem that we address is the use of CBMA data as prior in small underpowered fMRI studies. Based on some existing work on the topic, we develop a hierarchical model for fMRI studies that uses previous CBMA findings as a prior for the location of the effects. Finally, we discuss a classical problem of meta-analysis, the file drawer problem, where studies are suppressed from the literature because they fail to report any significant finding. We use truncated models to infer the total number of non-significant studies that are missing from a database. All our methods are tested on both simulated and real data.

# Chapter 1

# Introduction

Functional magnetic resonance imaging (fMRI) is a non-invasive, non-radioactive imaging technique that allows us to measure a person's brain activity while they perform a series of tasks. fMRI is based on a fundamental link between brain activity and blood flow: when there is a rise in neuronal activity in a region of the brain, there is also a local increase of blood flow in that region [Ogawa *et al.*, 1990]. This so-called haemodynamic response, which arises to meet the high demands for oxygen in the area, actually leads to a surplus of local blood oxygen whose magnetic susceptibility can then be detected by an MRI device [Kwong *et al.*, 1992]. Hence, one can easily establish a link between a stimulus and some brain region, just by studying the strength of the observed fMRI signal over time and throughout the brain: if a certain behaviour consistently induces a change in the fMRI signal in a certain brain region, then there is evidence that this region plays some role in the processing of the task.

This tool has motivated researchers to investigate the effect of several interesting tasks and thus led to an explosive growth in the use of fMRI as well as significant developments in our understanding of the human brain function [Raichle, 2003]. Some examples include the differences in brain function between maternal and romantic love [Bartels and Zeki, 2004], the effect of alcohol while performing simulated driving [Calhoun and Pearlson, 2012] or the effect of doing nothing at all [Cole *et al.*, 2010]. The availability of MRI scanners, inexpensive computational resources and accessible analysis software has made fMRI an ubiquitous tool in psychology, neurology and psychiatry, in addition to new areas like neuromarketing [Zurawicki, 2010] and neuroeconomics [Glimcher *et al.*, 2008].

Nevertheless, there are a variety of factors that limit the interpretability of fMRI results. The principal limitation is the small sample sizes that typify fMRI

studies [Carp, 2012]. For example, in a recent review of emotions Lindquist *et al.* [2012] found a median sample size of 11. Consequently, individual studies suffer from low power and hence low reproducibility [Button *et al.*, 2013], and it is unsurprising that the validity of fMRI is being challenged in both the scientific [Vul *et al.*, 2009] and popular [Shermer, 2008] literature (see Farah [2014] for an even-handed review).

Meta-analysis, the statistical process of combining the results of independently conducted studies to increase power and obtain findings that are more likely to generalise [Hedges and Olkin, 1985], provides a natural way to address the limitations of single experiments (which we explain in detail in the following chapter). Meta-analysis of functional neuroimaging data is indeed an active field of research [Wager *et al.*, 2009; Yarkoni *et al.*, 2010], whose growth is facilitated by the constantly increasing body of literature in fMRI but also the numerous challenges that arise due the high dimensionality of single experiment data. The main challenge lies in that instead of reporting the full outcome on an fMRI experiment, that is the 3D volumes of test statistics, authors generally only report the spatial coordinates of local maxima in significant regions of these images [Salimi-Khorshidi *et al.*, 2009]. As a result, the standard tools that are used in meta-analyses [Hedges and Olkin, 1985; Spiegelhalter *et al.*, 2004; Hartung *et al.*, 2008, among others] are no longer applicable and hence new tools need to be developed.

The objectives of this dissertation is to address the following still open problems:

- Review existing approaches for fMRI meta-analysis using spatial coordinates, explain the merits and disadvantages of these methods and identify the still open questions in the field.

- Develop a framework that enables meta-regression, the use of study characteristics as covariates in fMRI meta-analysis. Such a framework can be used in order to understand the impact that these covariates have in the outcome of a study and hence explain possible differences in reported results.

- Build a model for single fMRI studies that can make use of existing meta-analysis results as prior information. This is particularly useful for studies with few participants that are underpowered and so need the input from previous studies to be able to detect subtle effects.

- Estimate the file drawer quantity, the total number of studies that are missing from the fMRI literature because they have no significant findings. The file drawer problem is a major concern in classical meta-analysis as it can lead to

wrong estimates of the underlying effect and thus undermine the usefulness of meta-analysis.

The thesis is organised as follows. In Chapter 2 we review some background which is used in subsequent chapters. In the first part we sketch the typical fMRI experiment so the reader becomes familiar with the particularities of the data in hand. The second part provides some theory on spatial point processes, i.e. random sets of points in the $d$-dimensional Euclidian space, upon which our methods are built. In Chapter 3 we perform a literature review of coordinate-based meta-analysis methods. The existing tools are evaluated through simulation studies and real data analysis, followed by a discussion that emphasises on strengths and weaknesses of every approach.

Chapters 4 and 5 form the main body of the text. In the former, we develop a novel fMRI meta-analysis model based on log-Gaussian Cox processes [Møller *et al.*, 1998]. The model can provide useful inferences regarding meta-analysis data and improves upon existing methods in the sense that it can account for study characteristics in the analysis. In the latter, Chapter 5, we present a 3 level hierarchical model for single fMRI studies. Our method, built upon the work of Xu *et al.* [2009], can be viewed as an alternative to existing approaches with the extra advantage that it addresses the problem of using meta-analysis data as prior information in new fMRI studies. In both cases, we pay special attention to the computational tools that are used to enable inferences which are carried out under the Bayesian paradigm.

In Chapter 6 we propose a zero-truncated regression model to estimate the prevalence of non-significant fMRI studies that are suppressed from the literature. The model assumes that the entire population of studies, both significant and non-significant, arise from a common mechanism and hence uses traits of the observed units to infer upon the missing part. Finally, in Chapter 7 we summarise our contributions and set some directions for future research.

# Chapter 2

# Background

## 2.1 Neuroimaging background

What follows is a very brief review of fMRI and the practical steps involved in a fMRI study. For a more detailed introduction, see Lindquist [2008] for review of fMRI for statisticians, or Kim and Ogawa [2012] for a detailed, technical review of the meaning of the fMRI signal; Huettel *et al.* [2009] provide accessible textbook treatment, while Poldrack *et al.* [2011] give a practical, data-analysis-oriented perspective.

The objective of a single fMRI study is to identify the neural correlates of a physical, mental or perceptual process. When neurons in a region of the brain increase their firing rate, there is an increased demand for oxygen which is met by a localised increase in blood flow. The magnetic resonance signature, or susceptibility, of oxygenated and de-oxygenated blood differs, and thus a MRI scanner can capture changes in local oxygenation. This mechanism is known as the *Blood Oxygenation Level-Dependent* (BOLD) effect.

During an fMRI acquisition, participants lie flat in the scanner and are asked to perform a series of tasks, such as viewing images or reading texts, while the MRI scanner measures the BOLD signal. For each participant, the data takes the form of a time series of images, 3D snapshots of signal measurements all over the brain. The typical acquisition lasts 6-12 minutes, with data collected every 2 seconds, producing data on a grid with, typically, 2mm $\times$ 2mm spacing in-plane and 2mm-4mm slices, producing anywhere from 40,000 to over 100,000 voxels (volume elements) in the brain. Note that this is quite coarse spatial resolution, and separate, fine-resolution images (e.g. 1mm $\times$ 1mm $\times$ 1mm) are also taken to depict an individual's anatomy.

Before the raw data can be analysed, a series of preprocessing steps needs to be undertaken [Poldrack *et al.*, 2011]. These include motion correction, which ac-

counts for movements during the acquisition, and spatial smoothing which increases the signal-to-noise ratio. Smoothing is performed by convolution of the data with 3D gaussian kernels of diagonal covariance matrix. By convention, the kernel is not specified by its variance of standard deviation, but in terms of *full width at half maximum* (FWHM). For one dimension, e.g. $x$, the relationship between FWHM and standard deviation $\sigma_x$ is:

$$\text{FWHM}_x = \sigma_x \sqrt{8 \log 2}. \tag{2.1}$$

Isotropic kernel sizes of between 4 FWHM and 8 FWHM are common. To make data comparable across subjects, a crucial step is "spatial normalisation", the process of warping all subjects to a standard brain template, or brain atlas. There are different atlases available, but essentially all authors use either the Talairach atlas [Talairach and Tournoux, 1988] or the Montreal neurological institute (MNI) atlas (see Fig. 2.1).

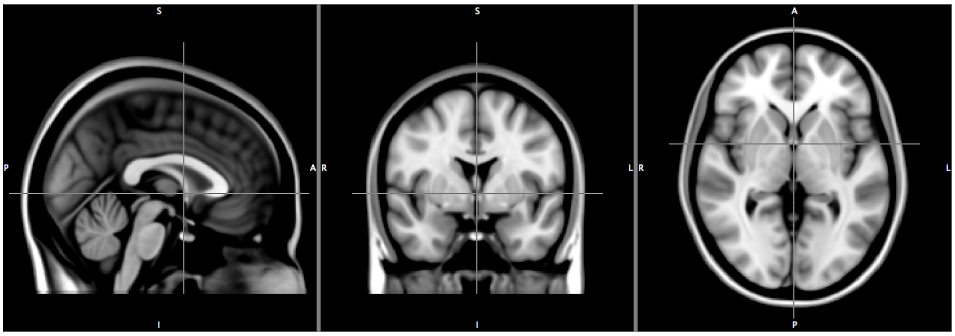



Figure 2.1: An average brain in MNI space. Note the directional labels at the edge of each panel: P for Posterior, A for Anterior, S for Superior, I for Inferior, L for Left and R for Right. The origin approximately corresponds to an anatomical structure known as the anterior commissure.

After spatial normalisation, all subjects' data exist in a common space. Specifically, we can assume that a given voxel corresponds to (roughly) the same region in all subjects' brains. Statistical analysis then proceeds in a massively univariate approach, by fitting a model at each voxel independently of every other voxel. Let $y_{ik}^v$ be the observed BOLD signal measurement for subject $i$ in voxel $v$ at time interval $k$, $k = 1, \ldots K$[1], $i = 1, \ldots, n$ and $v = 1, \ldots, V$. For subject $i$, the time series regression model at voxel $v$ is:

$$\mathbf{Y}_i^v = \mathbf{X}_i^v \boldsymbol{\beta}_i^v + \boldsymbol{\epsilon}_i^v, \tag{2.2}$$

[1]We have assumed a common time series length $K$ over subjects only for convenience

where $\mathbf{Y}_i^v = [y_{i1}^v, \ldots, y_{iK}^v]$ is time series of measurements at voxel $v$, $\mathbf{X}_i^v$ is the $K \times p$ design matrix representing $p$ experimental conditions (tasks) and $\boldsymbol{\epsilon}_i^v$ is the residual error. Due to the temporal correlation the error has some non-independent structure, $\boldsymbol{\epsilon}_i^v \sim \mathcal{N}_K\left(0, \mathbf{R}\,(\sigma_i^v)^2\right)$, where autocorrelation structure $\mathbf{R}$ expresses, for example, an autoregressive order 1 model. Generalised least squares are used to estimate model parameters $\hat{\boldsymbol{\beta}}_i^v$ [Mumford and Nichols, 2006] and hence the effect of a task for an individual can be tested with an appropriate *contrast* $\mathbf{c}$ meaning the estimated linear combination $\mathbf{c}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_i^v$ of parameter estimates that relates to the effect. Alternatively, one can combine the subject-specific regression coefficients in a "second level" model to test for the presence of a population effect [Mumford and Nichols, 2006].

In either case, the result is a 3D image of $T$ statistics (one per participant or one per study), with a value at each voxel in the brain that measures the evidence against the null hypothesis of no effect. The $T$ images are assessed either voxel-by-voxel, or by assessing the size of connected components, or *clusters*, after thresholding the $T$ image at an arbitrary threshold. See Friston *et al.* [2002], Mumford and Nichols [2006] and Mumford and Nichols [2009] for a detailed review of different approaches for the statistical analysis of fMRI data.

An essential issue in the statistical analysis of fMRI data is *multiple testing*. A $T$ statistic image can have 100,000 or more voxels in the brain, requiring 100,000 simultaneous tests for every contrast of interest. Under a global null hypothesis of no effect in any voxel, we therefore expect around 5,000 false positives using the classic significance level of $\alpha = 0.05$. In the early history of fMRI (roughly 1992-2002), arbitrary rule-of-thumb thresholding procedures were common, like a combination of an uncorrected voxel-wise $\alpha = 0.001$ and cluster size threshold $k \geq 10$ (only clusters of size 10 voxels or more). Thresholding methods that controlled the Familywise Error (FWE), the chance of one or more false positives, later became widespread using either Random Field Theory or permutation (see Nichols and Hayasaka [2003] for a review of FWE methods in neuroimaging). More recently, the *False Discovery Rate* (FDR), the expected proportion of false positives among positive findings [Benjamini and Hochberg, 1995] was introduced for the thresholding of $T$ images [Genovese *et al.*, 2002].

In any other discipline of science, publishing the point estimate, standard error, test statistic and $p$-value for an effect would be considered best practice, if not just the minimal information to report. In neuroimaging, each of these quantities is 3D image, and sharing such large data files was considered impractical 25 years ago when fMRI was first developed. Yet even today there is general resistance

towards sharing the full images. Instead, the only thing authors routinely report is the $x, y, z$ atlas coordinates of activation peaks. Going forward we will call these coordinates the *foci* (singular *focus*). In other words, the results of an fMRI study are summarised in a list of foci. Based on author preference and software defaults foci can either *singly* reported that is one focus per significant region or *multiply* reported that is two or more per significant region.

### 2.1.1 Limitations of individual studies and meta-analysis

There are three aspects of fMRI experiments that challenge the utility of individual studies. Firstly, fMRI studies suffer from low power. The typical sample size of an fMRI study is small, and the majority of experiments involves far less than 20 participants [Carp, 2012]. While power depends on the (unknown) true effect size, at least one empirical study supported the notion that fMRI $n$'s are too small. By sub-sampling from a large sample ($n = 150$), Thirion *et al.* [2007] found that analyses with 20 or fewer subjects were poor approximations of the full 150-subject result. Further, Type I error rates are likely to be high, especially for older papers that did not use inference procedures corrected for multiple testing. Using a survey of publications' thresholding methods, Wager *et al.* [2007] estimated that 17% of all reported foci are false positives. Finally, neuroimaging studies suffer from low test-retest reliability. For example, when scanning a group of subjects twice, once and then 7 days later, Raemaekers *et al.* [2007] found intra-class correlations for BOLD fMRI activations ranged from 0 to 0.88.

Apart from these inherent limitations, the way fMRI studies are carried out also exhibit great heterogeneity. Each step of a neuroimaging study can be implemented in various ways and there is no standard way to present a stimulus, preprocess the data or construct the linear model for the BOLD response. As a result, there is only a partial agreement in how experiments are conducted. For example, in an analysis of 241 fMRI studies Carp [2012] observed 223 different analytical strategies. Results heavily depend on the type of analysis employed [Button *et al.*, 2013], thus it is not uncommon to observe discrepancies in the outcomes of studies that investigate the same scientific question. Consequently, it is exceptionally hard to yield a conclusion. All reasons combined support the use of meta-analysis to account for these problems and draw more reliable inferences.

A well performed meta-analysis can tackle the aforementioned issues by modelling the observed heterogeneity between studies, combining the available information to increase power and ultimately separating the consistent findings from those that happened by chance. There exist two broad approaches for meta-analysis of

neuroimaging studies: *image-based meta-analysis* (IBMA), if the full $T$ statistic images are available, and *coordinate-based meta-analysis* (CBMA) if only foci are reported. IBMA proceeds by means of some common meta analytic tools applied to each voxel of the images along with either FWER or FDR corrections for multiple testing (see Hartung *et al.* [2008] for an overview of conventional meta analysis, and Lazar *et al.* [2002] for a review of IBMA methods). It is self-evident that the transition from full statistical images (100,000+ voxels) to the list of reported foci involves a heavy loss of information. In a comparative study, Salimi-Khorshidi *et al.* [2009] demonstrated the benefits of using IBMA over CBMA. However, the overwhelming majority of researchers rarely provide the full images, thus CBMA still constitutes the main approach for the meta analysis of fMRI data. We review existing CBMA methods in the following chapter.

## 2.2 Spatial point processes background

Spatial point processes are random sets of points in the $d$-dimensional Euclidian space, where both the number and location of points is random. Spatial point processes have been used in several areas of applications such as astronomy where points can represent locations of stars in a galaxy [Babu and Feigelson, 1996] or ecology where for example points can be the locations of ant nests in an region [Harkness and Isham, 1983]. In our applications we use spatial point processes to model reported peak activation foci that are reported by a task fMRI experiment.

The aim of this section is to give the necessary definitions and explain the basic spatial point process models that are used in subsequent chapters. Most of the results are taken from Møller and Waagepetersen [2004] to whom the reader can refer for a complete, theoretical treatment of the topic. Some other good references include Illian *et al.* [2008] and Diggle [2013] whereas Daley and Vere-Jones [2002] discuss point processes in general spaces.

We start with a spatial point process $\mathbf{X}$, a random countable subset of a space $\mathcal{B} \subseteq \mathbb{R}^d$. For us, $d = 3$ and $\mathcal{B}$ is the human brain. We use lower case letter $\mathbf{x}$ to refer to some realisation of $\mathbf{X}$ and require $\mathbf{x}$ to be locally finite subsets of $\mathcal{B}$, that is $n(\mathbf{x} \cap B) < \infty$ for any bounded $B \subseteq \mathcal{B}$, where $n(\cdot)$ stands for the cardinality of a set. Thus, the range of $\mathbf{X}$ is the space of locally finite point configurations $N_{\text{lf}}$ where:

$$N_{\text{lf}} = \{\mathbf{x} \subseteq \mathcal{B} : n(\mathbf{x} \cap B) < \infty, \forall \text{ bounded } B \subseteq \mathcal{B}\}.$$

The most basic spatial point process model is the *spatial Poisson process*. Poisson processes are used as a model class complete spatial randomness or no

interaction between points. For a point process $\mathbf{X}$ the total number of points in any $B \subseteq \mathcal{B}$, say $N_{\mathbf{X}}(B)$, is a random variable. The Poisson point process can be defined in terms of the intensity measure $\mu(\cdot)$, the expected number of points in any subset $B$ of its support:

$$\mu(B) = \mathbb{E}\left[N_{\mathbf{X}}(B)\right].$$

In particular, we say that $\mathbf{X}$ is a Poisson point process when the following properties are satisfied [Møller and Waagepetersen, 2003]:

- for any $B \in \mathcal{B}$ with $\mu(B) < \infty$, $N_{\mathbf{X}}(B) \sim \text{Pois}(\mu(B))$, the Poisson distribution with mean $\mu(B)$
- for any disjoint sets $B_1, \ldots, B_n \subseteq \mathcal{B}$ with $n \geq 2$, the random variables $N_{\mathbf{X}}(B_1), \ldots, N_{\mathbf{X}}(B_n)$ are independent.

Sometimes, for a Poisson process $\mathbf{X}$ on $\mathcal{B}$ we can write that:

$$\mu(B) = \int_B \lambda(\xi)\, d\xi,$$

for some non-negative function $\lambda$ for which $\int_B \lambda(\xi)\, d\xi < \infty$ for any bounded $B \in \mathcal{B}$. Then we say that $\mathbf{X}$ is a Poisson point process with *intensity function* $\lambda$ and denote $\mathbf{X} \sim \mathcal{PP}(\mathcal{B}, \lambda)$. When $\lambda$ is constant in $\mathcal{B}$ then $\mathbf{X}$ is homogenous; otherwise it is said to be inhomogenous. Let $\mathbf{X} \sim \mathcal{PP}(\mathcal{B}, \lambda)$. The density of this measure with respect to the measure induced by the $\mathcal{PP}(\mathcal{B}, 1)$ is [Møller and Waagepetersen, 2004]:

$$\pi(\mathbf{x} \mid \lambda) = \exp\left(|\mathcal{B}| - \int_{\mathcal{B}} \lambda(\xi)\, d\xi\right) \prod_{x \in \mathbf{x}} \lambda(x),$$

where $|\mathcal{B}|$ is the volume of the brain. The above equation can be used to estimate the intensity of a Poisson process $\mathbf{X}$ based on one or more realisations.

A natural extension of the spatial Poisson process is the spatial *Cox process*. The Cox process occurs when we allow the intensity function of a Poisson process to be a random variable itself. More precisely, let $\Lambda = \{\Lambda(\xi) : \xi \in \mathcal{B}\}$ be a non-negative random field. If $\mathbf{X} \mid \Lambda = \lambda \sim \mathcal{PP}(\mathcal{B}, \lambda)$ then we say that $\mathbf{X}$ is a spatial Cox process driven by random intensity $\Lambda$. The Cox process fits naturally into the Bayesian framework where we obtain a posterior distribution for an intensity function. Note that it is not possible to distinguish a Cox process from a Poisson process based on only one realisation.

Another interesting spatial point process model is the *independent cluster process*. This model is particularly useful when we observe aggregation of points

around some centers. For example, Waagepetersen and Schweder [2006] build a model for Minke whale locations in the Atlantic, based on the fact that these tend to cluster around areas with high prey density. The independent cluster process is obtained as following. Let $\mathbf{y}$ be the realisation of a point process $\mathbf{Y}$ on $\mathcal{B}$, the parent process. We associate with each $y \in \mathbf{y}$ a point process $\mathbf{X}_y$ of offsprings centered around $y$. If we further assume that $\mathbf{X}_y$ are independent one of another, then $\mathbf{X} = \bigcup_{y \in \mathbf{y}} \mathbf{X}_y$ is an independent cluster process [van Lieshout and Baddeley, 2002]. Typically, it is assumed $\mathbf{X}_y$ are Poisson processes with some intensity $\rho(\cdot \mid y)$; in this case the intensity of the point process $\mathbf{X}$ is $\lambda(\cdot \mid \mathbf{y}) = \sum_{y \in \mathbf{y}} \rho(\cdot \mid y)$.

Finally, it is sometimes the case that the points in a point pattern carry some extra information. Møller and Waagepetersen [2007] present a dataset consisting of the location of 134 Norwegian spruces and it is noted that for each tree the stem diameter is also recorded. This leads to a marked process. Consider a point process $\mathbf{X}$ on $\mathcal{B}$. Given some mark space $\mathcal{M}$, if we attach a random mark $m_x$ to each $x \in \mathbf{X}$, the process $\mathbf{Y}\{\{x, m_x\} : x \in \mathbf{X}\}$ is a marked process. If we further assume that marks are independent of the points and independent one of another then the density of $\mathbf{Y}$ can be written as $\pi(\mathbf{Y}) = \pi(\mathbf{X}) \prod_{x \in \mathbf{X}} \pi(m_x)$. Note that the mark space $\mathcal{M}$ can be very general; in the tree example above $\mathcal{M} = (0, \infty)$.

# Chapter 3

# The coordinate-based meta-analysis of fMRI data: a review

## 3.1 Introduction

The limitations of single experiments (see Section 2.1.1 for a discussion), along with the historical lack of data sharing, quickly presented researchers in the field of fMRI with a challenge. The standard meta-analytic tools used in other fields (see for example Hartung *et al.* [2008] for a fairly recent review) could not be applied to the coordinate data and hence there was a need for new methodologies. Early works mainly utilised exploratory data analysis and visualisation techniques to blend the results from different studies [Fox *et al.*, 1998] and it was not until the early 2000's that the first methods for CBMA were proposed [Fox *et al.*, 1997; Turkeltaub *et al.*, 2002; Nielsen and Hansen, 2002; Wager *et al.*, 2003]. Since then, many new methods and modifications appeared in the neuroimaging [Laird *et al.*, 2005; Wager *et al.*, 2007; Radua and Mataix-Cols, 2009; Turkeltaub *et al.*, 2012; Caspers *et al.*, 2014, to name a few] as well as the statistics [Kang *et al.*, 2011; Yue *et al.*, 2012; Kang *et al.*, 2014] literature.

All of these methods share the same goal: to identify areas of the human brain that show consistent activation across studies. The different approaches broadly fall into two main categories: *kernel-based* and *model-based* methods. In what follows, we present the most widely used methods in both categories. We start by setting the notation used throughout the chapter.

A typical CBMA dataset consists of a list of foci from $I$ independent studies.

Each study $i$, $i = 1, \ldots, I$ comes with a set of 3-dimensional coordinates $\mathbf{x}_{ik} \in \mathcal{B}$, where $\mathcal{B} \subset \mathbb{R}^3$ is the standard atlas space and $k$ indexes the multiple foci for a particular study. Table 3.1 is part of a real dataset from a meta-analysis of emotion studies that will be analysed for the purposes of this review. In this example, $\mathbf{x}_{52}$ would correspond to the second foci ($[-34, 52, 8]^{\mathrm{T}}$) in the fifth study (Baker 1997, emotion). Note that some of the studies (e.g. Damasio 2000, fear and Damasio 2000, anger) are obtained from the same experiment (publication); we treat these studies as independent following the standard conventions in the field. Finally, we will denote as $v = [v_x, v_y, v_z] \in \mathcal{B}$ the center location of a particular voxel in the brain atlas, $v = 1, \ldots, V$.

| **Author** | **Year** | **Emotion** | **X** | **Y** | **Z** | **Participants** |
|---|---|---|---|---|---|---|
| Damasio | 2000 | fear | -10 | -62 | -17 | 23 |
| | | | -1 | -66 | -1 | 23 |
| | | | 34 | 3 | 32 | 23 |
| Damasio | 2000 | anger | -2 | -29 | -12 | 23 |
| Philips | 2004 | disgust | 4 | -20 | 15 | 8 |
| | | | 7 | -17 | 9 | 8 |
| | | | 4 | -63 | 26 | 8 |
| Baker | 1997 | sad | 36 | 20 | -8 | 11 |
| | | | -44 | 32 | -8 | 11 |
| Baker | 1997 | happy | -26 | 28 | 0 | 11 |
| | | | -34 | 52 | 8 | 11 |
| Williams | 2005 | anger | 7 | 31 | 28 | 13 |
| | | | 7 | 28 | -7 | 13 |
| ... | ... | ... | ... | ... | ... | ... |

Table 3.1: A subset of data from a meta-analysis study of emotions.

## 3.2 CBMA methods

### 3.2.1 Kernel-based methods

The most widely used kernel-based methods are the *multilevel kernel density analysis* [Wager *et al.*, 2007, MKDA], the *activation likelihood estimation* [Eickhoff *et al.*, 2012, ALE] and the *signed differential mapping* [Radua *et al.*, 2012, SDM]. All these methods share the same rationale. Briefly, one starts by creating focus maps: that is, full brain images obtained by smoothing of each reported activation with a spatial kernel. Obviously, there are as many focus maps as the total number of foci. Secondly, the focus maps corresponding to a particular study are combined to create

the study-specific maps. These per-study images are subsequently combined into a single image that represents the evidence for consistent activation (clustering). Significance of these images is assessed with a Monte Carlo test under the null hypothesis of complete spatial randomness. We now discuss MKDA, ALE and SDM in detail.

**Multilevel kernel density analysis (MKDA)**

First introduced by Wager *et al.* [2003], MKDA was modified to its current version by Wager *et al.* [2007]. To obtain the focus maps, $M_{ik}$, one places a sphere of unit intensity and radius $r$ centred at each focus:

$$M_{ik}(v) = \mathbf{1}_{\{||v,\mathbf{x}_{ik}||\leq r\}}, \tag{3.1}$$

where $||\cdot,\cdot||$ stands for the Euclidian distance. The study specific images, $M_i$, are then obtained by applying the maximum operator to the focus maps of the study. The procedure can be expressed by the following formula:

$$M_i(v) = \begin{cases} 1, & \exists k \text{ s.t. } ||v,\mathbf{x}_{ik}|| \leq r \\ 0, & \text{otherwise} \end{cases}. \tag{3.2}$$

We call $M_i$ the *comparison indicator maps*. A value of 1 means that there is activation within distance $r$ of a given location. Wager *et al.* [2004] suggest giving $r$ a value of 10 or 15mm. The MKDA statistic image $m$ is given as a weighted combination of $M_i$:

$$m(v) = \frac{1}{\sum_i w_i} \sum_{i=1}^{I} w_i M_i(v). \tag{3.3}$$

The weights $w_i$ are usually chosen to be proportional to the number of participants in each study thus allowing for studies with larger sample size to contribute more to the value of the statistic. If the weights are all set to 1 then $m(v)$ denotes the proportion of studies that reported activation within distance $r$ to $v$. Large values of $m\,(v)$ suggest systematic clustering of foci around its location.

The distribution of the MKDA statistic does not have a closed form and thus Monte Carlo testing is used to assess significance. Several synthetic datasets are created by uniformly drawing peak locations from $\mathcal{B}$, keeping the original number of foci fixed. The $m$ statistic is calculated for these datasets and the maximum value is saved at each replicate. This produces a sample of the maximal statistic under the null hypothesis of random foci allocation. The sample is then used to obtain FWE

corrected $p$-values [Kober *et al.*, 2008] as suggested by Nichols and Holmes [2002]. Recently, Costafreda *et al.* [2009] derived a parametric significance test based on the properties of the spatial Poisson process. For applications of MKDA on real data see Etkin and Wager [2007] and Kober *et al.* [2008].

### Activation likelihood estimation (ALE)

The motivating idea behind ALE is to represent the uncertainty about the true location of a focus using a spatial Gaussian kernel [Turkeltaub *et al.*, 2002]. Let $L_{ik}$ be the map based on a single focus $\mathbf{x}_{ik}$,

$$L_{ik}(v) = c\phi_3(v \mid \mathbf{x}_{ik}, \sigma_i^2\mathbf{I}), \tag{3.4}$$

where $\phi_3(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of a three dimensional Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ evaluated at $\mathbf{x} \in \mathbb{R}^3$, $\mathbf{I}$ is the identity matrix, and $c$ is the normalising constant. The Gaussian kernel used for ALE is analogous to the uniform kernel used for MKDA, but assigns higher values to the voxels closer to the foci. To help determine $\sigma_i$, Eickhoff *et al.* [2009] created a mapping between the number of participants in each study, $n_i$, and the standard deviation $\sigma_i$; however, the mapping is based on an empirical study consisting of 21 subjects and may be unsuitable for experimental paradigms other than the one used by the authors.

Based on the Gaussian assumption, $L_{ik}(v)$ represents the probability of $v$ being the true location of $\mathbf{x}_{ik}$. These maps are combined into a *modelled activation map*, $L_i(v)$, giving the probability that the closest focus is truly located at $v$ [Turkeltaub *et al.*, 2012]:

$$L_i(v) = \max_k L_{ik}(v), \tag{3.5}$$

under the assumption that foci in study $i$ are independent. The ALE statistic $\ell$ is then computed as:

$$\ell(v) = 1 - \prod_{i=1}^{I} \left(1 - L_i(v)\right). \tag{3.6}$$

Expression (3.6) was originally adopted by Turkeltaub *et al.* [2002] to quantify the probability that at least one activation occurs in voxel $v$; nevertheless, $\ell$ should not be viewed as a probability distribution since that would require the foci of a study to be independent of one another.

The Monte Carlo significance test of ALE is equivalent but slightly different to the one of MKDA. In particular, Eickhoff *et al.* [2009] observe that one can

directly derive null ALE values by recomputing $\ell$ with random voxel location, i.e.:

$$\ell^* = 1 - \prod_i (1 - L_i(v^*)), \tag{3.7}$$

where $L_i(v^*)$ are randomly selected voxels from the corresponding study map. Therefore, creating new datasets under the null hypothesis as in MKDA is not necessary. Recently, Eickhoff *et al.* [2012] showed that by using (3.7), it is possible to enumerate exhaustively all the possible outcomes, thus ensuring that the tail of the Monte Carlo distribution (where the inference is based) is better approximated. Thresholding to assess significance can be done by controlling the FDR or the FWE [Laird *et al.*, 2005] or by inferring on clusters rather individual voxels [Eickhoff *et al.*, 2012]. ALE has been used for several analyses including Delvecchio *et al.* [2012] and Konova *et al.* [2013].

**Signed differential mapping (SDM)**

SDM [Radua and Mataix-Cols, 2009] is a relatively new method that borrows several characteristics from both MKDA and ALE. The novelty of the method lies in incorporating the $T$ statistic values (when available) from the original studies. To make this point clear, imagine that a study investigates brain activation caused by a given task; in some regions of the brain hyperactivation will be observed while in others there will be underactivation. In both cases, significant values of the $T$ statistic will be recorded; these values will be large and positive in the first case and large and negative values in the second. This case is particularly interesting when difference in activation between tasks is being investigated.

Assume that $T_{ik}$ is reported $T$ value for the focus $\mathbf{x}_{ik}$. SDM will generate the focus maps $S_{ik}$ as:

$$S_{ik}(v) = \text{sign}\,(T_{ik}) \exp\left(-\frac{||v, \mathbf{x}_{ik}||^2}{2\sigma^2}\right). \tag{3.8}$$

Observe that a Gaussian kernel is used, exactly as in ALE. Authors suggest using a standard deviation $\sigma$ of approximately 10mm. The study maps are then:

$$S_i(v) = \begin{cases} -S_{\max}, & \sum_k S_{ik}(v) \leq -S_{\max} \\ \sum_k S_{ik}(v), & -S_{\max} \leq \sum_k S_{ik}(v) \leq S_{\max} \\ S_{\max}, & \sum_k S_{ik}(v) \geq S_{\max} \end{cases}, \tag{3.9}$$

where $S_{\max} > 0$ is the fixed, maximum allowed value. That is, the study map is

obtained as the sum of the corresponding focus maps, but is forced to lie within the interval $[-S_{\max}, S_{\max}]$ in the same way the MKDA study maps $M_i$ are given a maximum value of 1. Finally, the SDM statistic image, $s$, is calculated as the weighted mean of the study specific maps at each voxel:

$$s(v) = \frac{1}{\sum_i w_i} \sum_i w_i S_i(v). \tag{3.10}$$

Weights are once again proportional to number of participants in the study. Since the method averages both positive and negative findings, voxels that show contradicting results will not appear as significant. Inference is based on the same Monte Carlo scheme of MKDA and thresholding is done either by setting a highly conservative rejection point ($p < 0.001$) or controlling the FDR [Radua and Mataix-Cols, 2009].

Recently, a new version of the algorithm was proposed, in which the authors used the $T_{ik}$ values to reconstruct the original $T$ statistic images. That way, it is possible to incorporate both CBMA and IBMA data in the same analysis. For more details, see Radua *et al.* [2012]. The last contribution made on SDM lies in the use of anisotropic kernels in the analysis [Radua *et al.*, 2014]. Anisotropy can be easily incorporated in MKDA and ALE but its superiority to the current practice of using isotropic kernels is only based on empirical findings and thus should be further investigated. Published work utilising SDM for the analyses includes Richlan *et al.* [2011] and Fusar-Poli [2012].

### 3.2.2 Model-based Methods

Recently there has been growing interest in the development of model-based methodologies to address some of the limitations of kernel-based methods. These methods use ideas from spatial statistics to develop stochastic models for the analysis of foci. Unfortunately the literature on model-based methods is still very limited thus our review will be almost exhaustive. In particular, we will outline the *Bayesian hierarchical cluster process* model of Kang *et al.* [2011, BHICP], the *spatial binary regression* model of Yue *et al.* [2012] and the *hierarchical Poisson/Gamma random field* model of Kang *et al.* [2014, HPGRF]. In all cases, analyses are performed under the Bayesian paradigm and thus inferences are based on posterior distributions for each model's parameters.

### A Bayesian hierarchical independent cluster process model (BHICP)

Kang *et al.* [2011] proposed a hierarchical model based on an independent cluster process to describe the mechanism generating the foci. The model is structured into 3 levels, of which the lowest level, level 1, contains the observations (foci), while higher levels describe the study and population structure respectively. The distinction between singly and multiply reported foci is incorporated into the model. In the outline of the model below, we occasionally suppress the $k$ index so that $\mathbf{x}_i$ is the set of foci reported in study $i$, i.e. $\mathbf{x}_i = \bigcup_k \mathbf{x}_{ik}$.

Figure 3.1 provides a graphical representation of the model. At level 1, we have foci (Fig. 3.1 bottom, coloured circles). We denote with $\mathbf{X}_i$ for the underlying process generating the observations $\mathbf{x}_i$ in each study. As discussed in Section 2.1, we can have both multiply and singly reported foci. Thus, $\mathbf{X}_i$ consists of two mechanisms, one generating the multiply reported foci (Fig. 3.1 bottom, red circles) and one that is giving the singly reported foci (Fig. 3.1 bottom, green circles): $\mathbf{X}_i = \mathbf{X}_i^1 \cup \mathbf{X}_i^0$.

Multiply reported foci $\mathbf{X}_i^1$ can be viewed as an independent cluster process of points $\mathbf{X}_{i\xi}^1$, which are normally distributed around study activation centers $\xi \in \mathbf{y}_i$ with covariance matrix $\mathbf{\Psi}_\xi$ (a random mark attached to every $\xi \in \mathbf{y}_i$). That is, $\mathbf{X}_i^1 = \bigcup_{\xi \in \mathbf{y}_i} \mathbf{X}_{i\xi}^1$ and for all $\xi \in \mathbf{y}_i$:

$$\mathbf{X}_{i\xi}^1 \sim \mathcal{N}\left(\xi, \mathbf{\Psi}_\xi\right). \tag{3.11}$$

Singly reported foci, $\mathbf{X}_i^0$, come directly from the population activation centers $\mathbf{z}$ as realisations of $\mathbf{X}_{i\zeta}^0$, which are normally distributed around population centers $\zeta \in \mathbf{z}$ with covariance matrix $\mathbf{\Sigma}_\zeta$:

$$\mathbf{X}_{i\zeta}^0 \sim \mathcal{N}\left(\zeta, \mathbf{\Sigma}_\zeta\right). \tag{3.12}$$

To add more flexibility, the model allows for some singly reported foci to not cluster around any population center, say $\mathbf{x}_{i\emptyset}$. These foci are assumed to arise from a Poisson process $\mathbf{X}_{i\emptyset}$ of constant intensity $\epsilon_1$:

$$\mathbf{X}_{i\emptyset} \mid \epsilon_1 \sim \mathcal{PP}\left(\mathcal{B}, \epsilon_1\right). \tag{3.13}$$

Overall, $\mathbf{X}_i^0 = \left(\bigcup_{\zeta \in \mathbf{z}} \mathbf{X}_{i\zeta}^0\right) \cup \mathbf{X}_{i\emptyset}$ are the singly reported foci of a study.

At level 2, we have the unobserved study activation centers $\mathbf{y}_i$, which are the locations around which the multiply reported foci of a study cluster. The $\mathbf{y}_i$ are realisations of a point process $\mathbf{Y}_i$ and may either cluster around the population centres $\mathbf{z}$ (Fig. 3.1 middle, squares) or appear in random locations across the brain

(Fig. 3.1 middle, triangles). To account for the former, clustered study centers, $\mathbf{Y}_{i\zeta}$ are introduced as sets of points normally distributed around population centers $\zeta \in \mathbf{z}$ with variance matrix $\mathbf{\Sigma}_\zeta$. As for the latter, noise study centers are modelled as a homogenous Poisson process $\mathbf{Y}_{i\emptyset}$ with intensity $\epsilon_2$. Overall, $\mathbf{Y}_i = \left(\bigcup_{\zeta \in \mathbf{z}} \mathbf{Y}_{i\zeta}\right) \cup \mathbf{Y}_{i\emptyset}$, where:

$$\mathbf{Y}_{i\zeta} \sim \mathcal{N}\left(\zeta, \mathbf{\Sigma}_\zeta\right), \tag{3.14}$$

and

$$\mathbf{Y}_{i\emptyset} \mid \epsilon_2 \sim \mathcal{PP}\left(\mathcal{B}, \epsilon_2\right). \tag{3.15}$$

At the highest level (level 3), we have the population activation centres (Fig. 3.1 top, gray crosses). These are unobserved realisations $\mathbf{z}$ of an *a priori* homogenous Poisson process $\mathbf{Z}$ of intensity $\epsilon_3$:

$$\mathbf{Z} \mid \epsilon_3 \sim \mathcal{PP}\left(\mathcal{B}, \epsilon_3\right). \tag{3.16}$$

Population activation centres are the locations around which study activation centers and singly reported foci scatter. As such, they can be viewed as locations in the brain where an overall population effect exists.

The BHICP can be viewed as a random effects model as it allows for both within-study and between-study variability. Samples from the posterior distributions are obtained via MCMC. Several interesting quantities can be inferred upon such as regions of consistent activations (through the distribution of populations centers), the uncertainty in the location of study centers around the population centers (through $\mathbf{\Sigma}_\zeta$) and the variability of the foci within studies (through $\mathbf{\Psi}_\xi$).

### A Bayesian nonparametric binary regression model

Yue *et al.* [2012] use spatial logistic regression for a meta-analysis of emotion studies. For study $i$ and voxel $v$, let $y_i\left(v\right)$ be the binary outcome defined as:

$$y_i(v) = \begin{cases} 1 & \text{at least one focus at voxel } v \\ 0 & \text{no foci at voxel } v \end{cases}. \tag{3.17}$$

Note that the binary study images $\{y_i\left(v\right)\}_{v=1}^V$ are identical to the MKDA study maps $M_i\left(v\right)$. Logistic regression can be used to model the probability that a voxel is reported as a focus, $p_i(v) = \mathbb{P}\left(y_i\left(v\right) = 1\right)$. It is assumed that:

$$p_i(v) = H\left(z(v)\right), \tag{3.18}$$

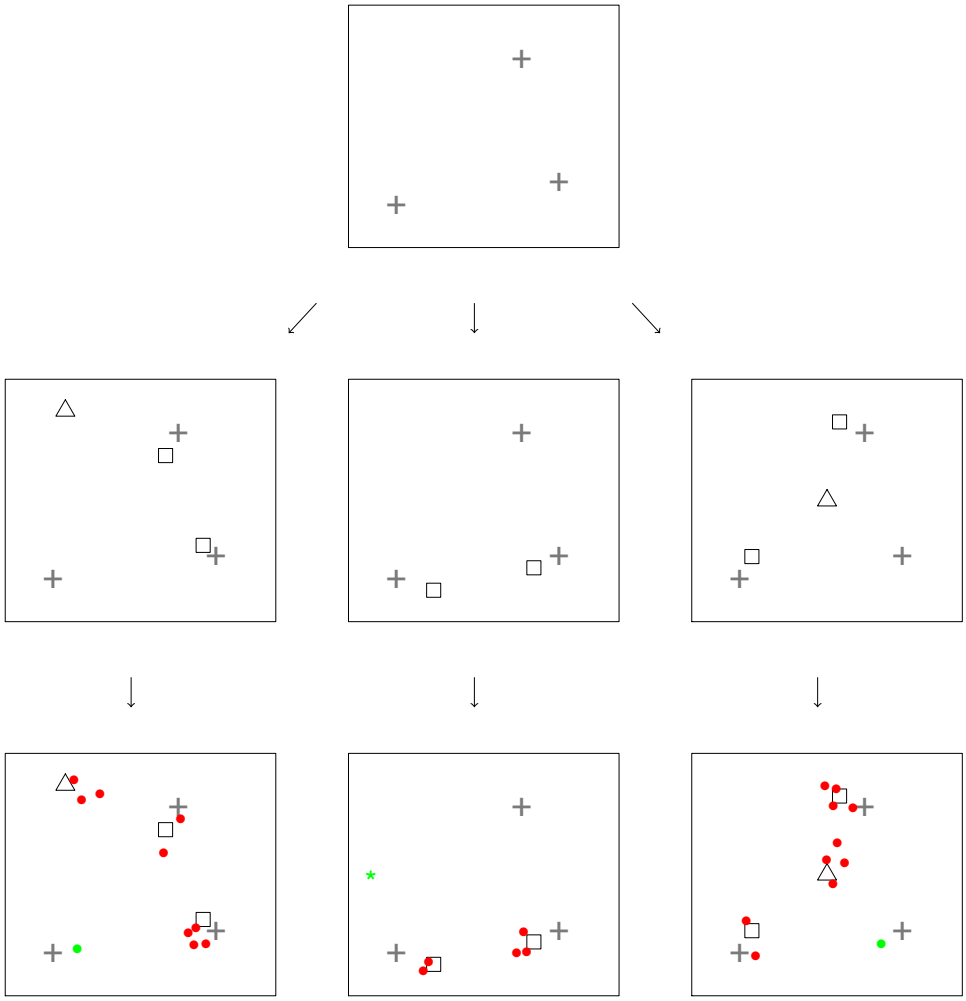


Figure 3.1: Realisation of the BHICP model for 3 studies. At level 1 (top) latent population centres (grey, $\mathbf{z}$) lie. At level 2 (middle) we have centres of multiply reported foci (black). These come either directly from population centres (squares, $\mathbf{y}_i$) or from background noise (triangles, $\mathbf{y}_{i\emptyset}$). Level 1 (bottom) contains the data ($\mathbf{x}_i$). These are multiply (red, $\mathbf{x}_i^1$) or singly (green, $\mathbf{x}_i^0$) reported foci. Singly reported foci come either directly from population centres (dots, $\mathbf{x}_{i\xi}^0$) or from a background Poisson process (asterisks, $\mathbf{x}_{i\emptyset}^0$).

where $H(\cdot)$ is the link function. The authors use the standard probit and logit link functions.

Spatial correlation is induced through the prior on $\{z(v)\}_{v=1}^{V}$. In particular, we assume that the process $z(v)$ is an adaptive Gaussian Markov random field [Yue and Speckman, 2010, aGMRF]. The aGMRF model defines the conditional distribution of $z(v)$ through a specific dependence with neighbouring voxels. A significant merit of the method is the inclusion of a local smoothness parameter $\gamma(v)$ for the aGMRF. This allows the method to automatically choose the amount of smoothing required depending on the amount of information available.

Authors further introduce a process $\psi_i(v)$, an indicator of whether the outcome variable $y_i(v)$ is miscoded; the case $\psi_i(v) = 1$ can either refer to both false positives, voxels that were falsely found as activated, and false negatives, voxels that were not reported as foci even though they were activated. The process $\psi(v)$ is not observed and hence is estimated along with the remaining model parameters.

Posterior probabilities of activation at each voxel are obtained through an auxiliary variable MCMC algorithm. Voxels with high posterior probabilities of being reported as foci are more likely to show an effect. A potential drawback of the method is that it can be currently applied only in two dimensions. In three dimensions, the value of one of the axes is held fixed, for example $z = c$, while the model is fitted for all available observations of the form $\mathbf{x}_{ik} = [(x_{ik})_1, (x_{ik})_2, c]$. Authors however, maintain that extending the model to three dimensions is possible.

### A hierarchical Poisson/Gamma random field model (HPGRF)

A neuroimaging meta-analysis will typically consider several subtypes of tasks. For example, a meta-analysis of emotion may classify the studies according to experiments on "happiness", "sadness", "pain", etc. Yet, the methods described previously are for a single homogeneous group of studies. Kang *et al.* [2014] propose a model that models each type of foci separately, allowing simultaneously for dependence between the $J$ different types.

Let $\mathbf{x}_{ij}$ be the set of foci reported by study $i$ for task type $j$. Suppose that $\mathbf{x}_{ij}$ are realisations of a Cox Process $\mathbf{X}_j$ driven by a random intensity measure $\mathbf{\Lambda}_j(d\xi)$. Conditional on $\mathbf{\Lambda}_j(d\xi)$, $\mathbf{X}_j$ are Poisson processes on the brain $\mathcal{B}$:

$$\mathbf{X}_j \mid \mathbf{\Lambda}_j(d\xi) \sim \mathcal{PP}(\mathcal{B}, \mathbf{\Lambda}_j(d\xi)). \tag{3.19}$$

In other words, we have $J$ underlying Cox processes, each one contributing a specific type of foci in some/all of the studies. The intensity measures $\mathbf{\Lambda}_j(d\xi)$ arise

from a convolution of a finite kernel measure $\mathbf{K}_j(d\xi, \zeta)$ and a Gamma Random Field $\mathbf{G}_j(d\zeta)$:

$$\mathbf{\Lambda}_j(d\xi) = \int_{\mathcal{B}} \mathbf{K}_j(d\xi, \zeta)\mathbf{G}_j(d\zeta). \tag{3.20}$$

The model arising from (3.19)-(3.20) is similar to the Poisson/Gamma random field model of Wolpert and Ickstadt [1998], who first introduced the idea of convolving a Gamma random field with a Poisson process. To introduce dependence between the different tasks, it is assumed that $\mathbf{G}_j(d\zeta)$ are independent realisations of a Gamma random field with common shape measure $\mathbf{G}_0(d\zeta)$ and inverse scale parameter $\beta$:

$$\mathbf{G}_j(d\zeta) \sim \mathcal{GRF}(\mathbf{G}_0(d\zeta), \beta). \tag{3.21}$$

Again, $\mathbf{G}_0(d\zeta)$ is a Gamma random field:

$$\mathbf{G}_0(d\zeta) \sim \mathcal{GRF}(\alpha(d\zeta), \beta_0). \tag{3.22}$$

An MCMC scheme is used for posterior computation. The HPGRF model allows for the detection of overall effects based on the posterior intensity $\mathbf{G}_0(d\zeta)$ or task-specific effects based on $\mathbf{\Lambda}_j(d\xi)$. Inference on types with fewer observations can be done by borrowing information from the remaining types through correlation under the common base intensity $\mathbf{G}_0(d\zeta)$. A significant benefit of the model is that it requires the specification of very few hyperparameters.

## 3.3 Evaluation of existing methods

One of the aims of this dissertation is to evaluate CBMA methods. A head-to-head comparison of existing methodologies is unfeasible, because the statistics described earlier have very different interpretations. Instead, we examine some characteristics of CBMA methods that show the drawbacks and merits of each. In what follows, we focus on the comparison between kernel-based and model-based methods. In Section 3.3.1 we conduct a series of simulations to study the sensitivity properties of the ALE algorithm that we think characterise other kernel-based methods as well. In Section 3.3.2, we apply the methods for which available software exists on a real dataset and compare the outputs. Finally, in Section 3.3.3 we move to a discussion.

### 3.3.1 ALE simulation study

Even though kernel-based methods have been extensively used for the analysis of neuroimaging data, their power properties have not been investigated on synthetic

datasets. We perform a simulation study to assess the power properties of the ALE method. In particular, we want to assess how the power of the algorithm evolves with respect to the number of studies in the meta-analysis and whether the method is robust to the inclusion of low quality studies. We choose ALE for three main reasons. Firstly, ALE is currently the most broadly used method for CBMA (based on a PubMed search for ALE, MKDA and SDM). Secondly, a recent review of kernel-based methods [Radua and Mataix-Cols, 2012] reported that the three kernel-based methods provide qualitatively similar results, thus we expect that our findings are indicative of MKDA and SDM methods as well. Finally, we strongly believe that the current version of ALE [Eickhoff *et al.*, 2012] provides the best approximation to the Monte Carlo test null distribution upon which inference is based.

We create meta-analytic datasets based on the following setup. Each simulated dataset consists of $I$ studies; of these, $Ip$ are valid while the rest $I(1-p)$ are noise, $0 \leq p \leq 1$. For the valid studies, we assume there exist 8 population centers around which foci cluster. A valid study detects each population center independently with probability 0.8. Conditional on detection, a study will report a singly reported focus with probability 0.4, two multiply reported foci with probability 0.35 or three multiply reported foci with probability 0.25. The foci are drawn from a three dimensional Gaussian distribution centered at the corresponding population center. As for the noise studies, we simply sample foci uniformly from the brain mask. The expected number of foci for both valid and noise studies is set to 13, similar to what we found in one of our applications (see Section 4.5).

We use study numbers of $I$ of $20, 40, 60, 80, 100$ and 120. For a given $I$ we successively set $p = 0, 0.05, 0.10, 0.15, ..., 0.95, 1$. For each distinct combination of $I$ and $p$ we create $B = 1,000$ datasets as described above, and apply the ALE algorithm [Eickhoff *et al.*, 2012] to each dataset. We use an $\alpha = 0.05$ FDR corrected threshold to access significance of the ALE statistic images. The following power-related quantities are recorded: 1) the probability that at least one of the 8 population centers is detected; 2) the probability all 8 centers are detected; 3) the mean number of centers detected in 1,000 runs; 4) the mean voxel-wise true positive rate, where "truly" active voxels are defined by the 95% probability spheres around the population centers.

Our findings are summarised in Figure 3.2 where quantities $1-4$ are plotted against the proportion of valid studies. One can observe that all 4 power measures increase monotonically to their maximal values of $1, 1, 8$ and 1, respectively, as the number of studies grows. For a given proportion of valid studies, we observe that the bigger the total sample size is, the higher the power. For a fixed sample size, the

power increases with the proportion of valid studies. Therefore, ALE is a consistent test. In Figure 3.3 we plot quantities $1-4$ versus the total number of valid studies, that is, $Ip$ instead of $p$. We see that the curves for different $I$ tend to coincide. This is a key robustness property of the ALE algorithm: that is, adding pure-noise studies does not degrade power detection.

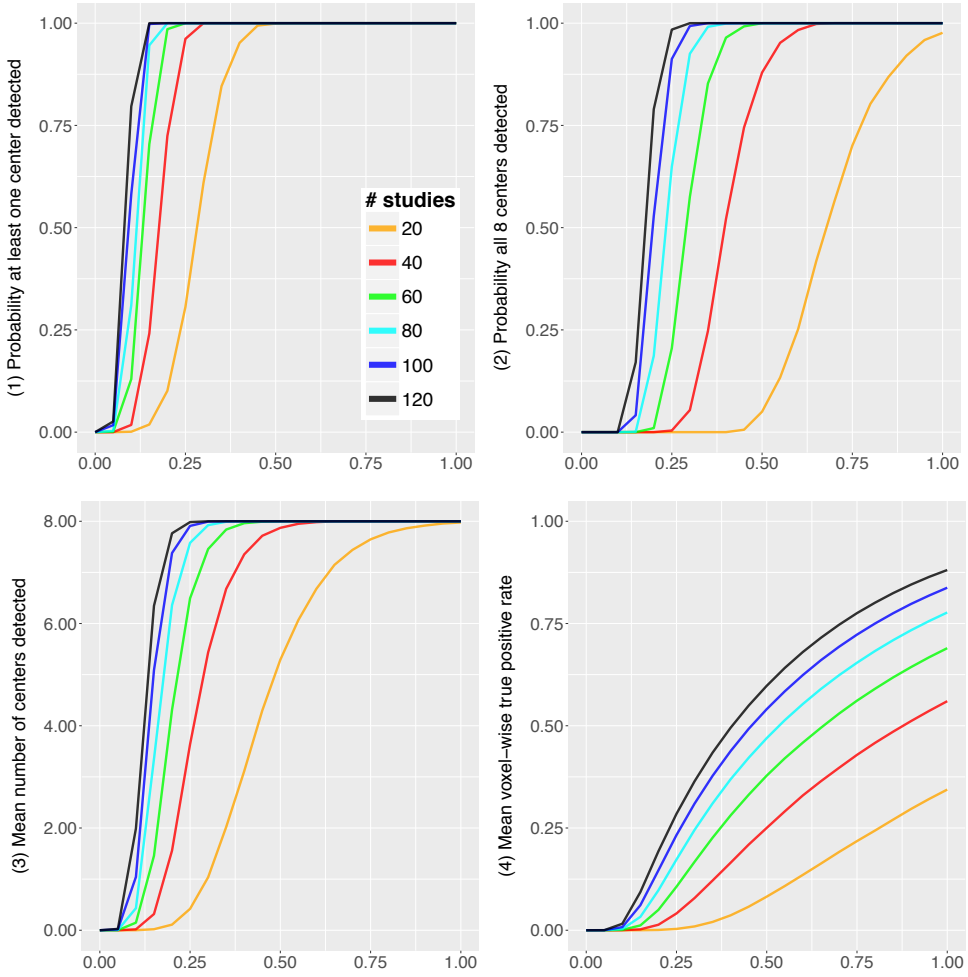



Figure 3.2: Results of the simulation study. Power properties of the ALE algorithm are plotted against the proportion of valid studies $p$. Top left: probability at least one center detected. Top right: probability all 8 centers detected. Bottom left: mean number of centers detected. Bottom right: mean voxel-wise true positive rate.

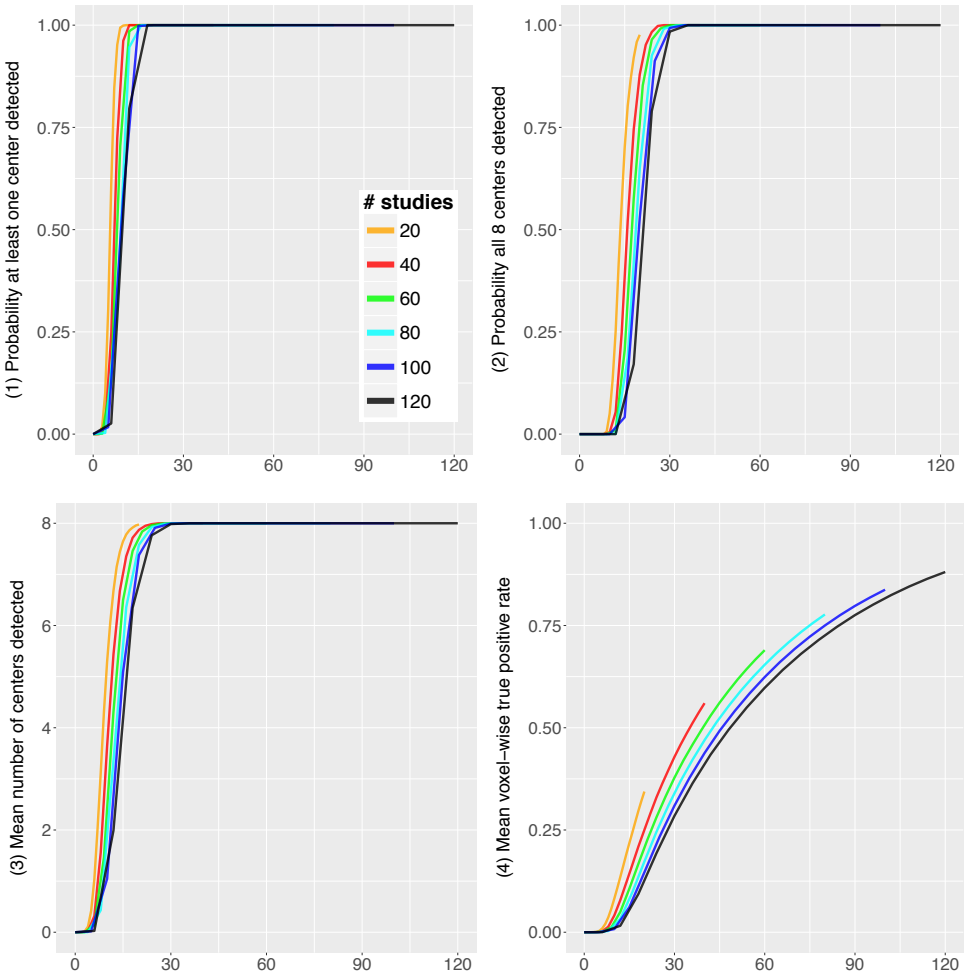



Figure 3.3: Results of the simulation study. Power properties of the ALE algorithm are plotted against the total number of valid studies $Ip$. Top left: probability at least one center detected. Top right: probability all 8 centers detected. Bottom left: mean number of centers detected. Bottom right: mean voxel-wise true positive rate.

### 3.3.2 Analysis of a real dataset

In this section, we perform a meta analysis of emotion studies that will facilitate the discussion of the next section. The dataset consists of 164 experiments conducted between 1993 and 2005. Eight emotion types appear in the dataset: affective, anger, disgust, fear, happy, mixed, sad and surprise. A total number of 2478 foci is reported with a mean value of foci per study close to 6. The goal of the analysis is to find regions of consistent activation across emotions. Due to the lack of software availability we only apply MKDA[1], ALE[2], SDM[3] and the BHICP[4]. Since those methods can not account for different task types we treat different emotions within an experiment as independent; this results into a total sample size of $I = 437$ studies (contrasts). The same dataset was analysed by Kober *et al.* [2008], Kang *et al.* [2011] and Yue *et al.* [2012].

The simulation parameters are set as following. For MKDA, we use a kernel size of $r = 10$mm, which is also the software default. A total of 10,000 Monte Carlo datasets are generated under the null hypothesis and used to the threshold the MKDA statistic image $m(v)$ at $\alpha = 0.05$, FWE corrected. ALE automatically assigns a kernel size for each study based on the total number of participants and uses the method of Eickhoff *et al.* [2012] to calculate the distribution of the statistic under the null hypothesis. The significance of the statistic image $\ell(v)$ is accessed with an FDR corrected $\alpha = 0.05$ threshold. For SDM we use an isotropic kernel of 20mm since it is the software default and do 500 Monte Carlo randomisations. For the BHICP we run the MCMC for 120,000 iterations saving once every 100 iterations. This results to a total sample size of 1,200 posterior draws, of which we discard the first 200 a burn-in. We use the same hyperparameter values as in Kang *et al.* [2011]. We now summarise the results.

Figure 3.4 shows statistic images obtained from the four methods considered, conditional on several values of the $z$ dimension. Note that for the BHICP we show only the study activation process intensity function. We see that all of the methods provide qualitatively similar results. More specifically, the regions of the brain that are mostly engaged in emotion processing are the right and left amygdala (Fig. 3.4, top and middle row). This finding is consistent with previous analyses of the same dataset [Kober *et al.*, 2008; Kang *et al.*, 2011; Yue *et al.*, 2012] as well as results of previous studies [Phelps and LeDoux, 2005; Costafreda *et al.*, 2008]. Other regions

[1] `http://wagerlab.colorado.edu/files/tools/meta-analysis.html`
[2] `http://www.brainmap.org/ale/`
[3] `http://www.sdmproject.com/software/`
[4] `http://www-personal.umich.edu/~jiankang/software.html`

where large statistic values are observed are the right and left cerebral cortex (Fig. 3.4, bottom row).

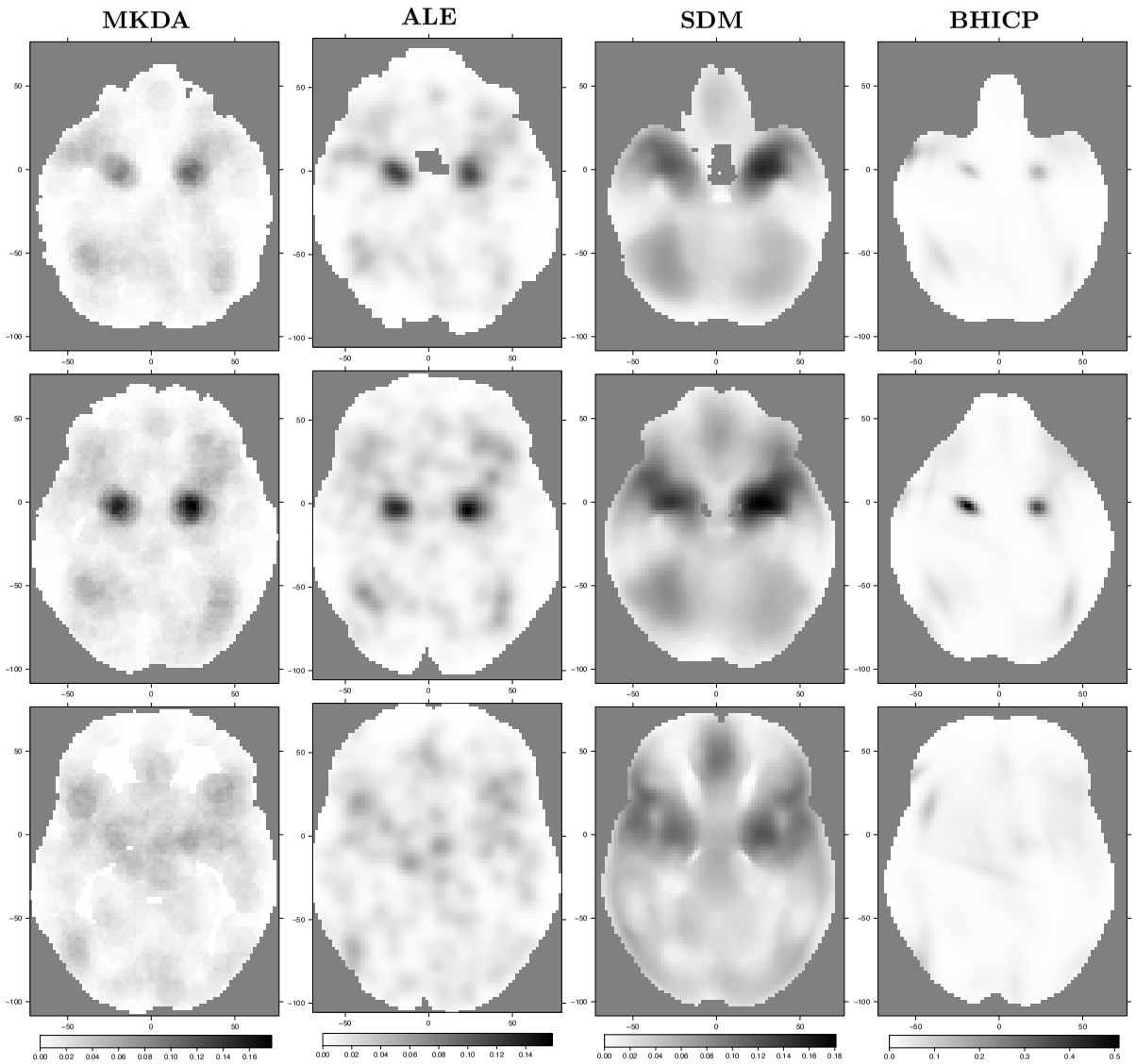



Figure 3.4: Qualitative comparison of methods in regions of high clustering of foci. Column 1 is MKDA, column 2 is ALE, column 3 is SDM and column 4 is the study activation center intensity. Rows 1-3 correspond to axial slice $z = -22$, $z = -16$ and $z = -2$ respectively.

### 3.3.3 Discussion

The meta-analysis of emotions suggests that the outputs obtained from model-based methods are qualitatively similar to those obtained from kernel-based methods. Nevertheless, the two approaches are fundamentally different.

In terms of computational time, it is clear that kernel-based methods outperform model-based methods (with the exception of SDM). For the emotion dataset used in the illustration of Section 3.3.2 ALE required approximately 15 minutes to run and MKDA required around 3 hours for 10,000 Monte Carlo replications. On the contrary, the BHICP model took roughly 16 hours for 120,000 MCMC iterations, Kang *et al.* [2014] needed 20 hours to complete the analysis and it is yet not possible to run the spatial binary regression model on the full brain.

Apart from running time, one needs to consider the ease of implementation. On the one hand, software for kernel-based methods can be applied to any dataset and will produce a pair of brain images: one with the value of the statistic and one containing the corresponding $p$-values. That results into a very automated procedure: one can directly compare the two images and see which voxels where significant and which were not. On the other hand, it is not straightforward to implement an MCMC scheme for one of the model-based methods. Prior specifications that are suitable for one dataset may be completely inappropriate for another. Further, it is not possible to know in advance how many iterations are required for the MCMC algorithm to converge, and convergence needs to be assessed as well. Overall, it may not seem appealing for a neuroimaging practitioner to adopt model-based methods and this explains why kernel-based methods are generally preferred in the outstanding majority of the analyses.

Nevertheless, despite the ease of implementation kernel-based methods allow for very limited inferences. First of all, kernel-based methods cannot be viewed as spatial models but are instead *massively univariate approaches* (MUA). This means that any inference is done voxel-by-voxel a no argument can be done for *a priori* defined groups of voxels or the entire brain. As an example, with kernel-based methods it is not possible to infer the total number of foci in a study; in fact, analyses condition on the total number of foci in a study. For model-based methods based on point processes such as the BHICP, the expected number of foci can be obtained by integrating the intensities over the brain. Further, MUA do not account for the spatial correlation and hence there is no borrowing of information across voxels. This leads to poor spatial precision as can be seen in Figure 3.4 where the BHICP statistic is more concentrated compared to the kernel-based methods.

Another important limitation of kernel-based methods is that there are no

standard errors given and therefore it is not possible to quantify the uncertainty for the effect estimate through confidence intervals, which in turn may result to misleading conclusions. For example, in Section 3.3.1 we found that power properties of ALE do not degrade with the inclusion of poor quality studies (see Fig. 3.3). However, since inferences remain unchanged, it is not possible to distinguish between cases with strong signal (few poor quality studies) and weak signal (many poor quality studies). Note that this is a fixed effects model property where a small proportion of the data drives the inference. Model-based methods tackle this through the standard errors obtained directly from the posterior distribution of any parameters of interest; when the signal is strong there is small variability in the posterior whereas when the signal is weak the variability is higher. Finally, kernel-based methods provide no adequate justification for the choice of kernel size parameter ($r$ for MKDA and $\sigma$ for ALE, SDM). Typically, its value is specified based on previous studies rather than being estimated from the data, and it remains constant across the brain regardless of the amount of smoothing required in each region. A bad choice of the kernel size can potentially affect the results, though. For example, in Figure 3.4 we see that in the statistic image of SDM the clusters appear to be bigger, because we have used a larger kernel size compared to the other kernel-based methods. Yue *et al.* [2012] automatically choose the amount of smoothing require introducing an extra smoothness parameters to their aGMRF.

Several other quantities of interest can be obtained from model-based methods. For the BHICP model it is possible to derive $(1-\alpha)\%$ credible ellipses for both population and study activation centers thus returning an estimate of within-study and between study-variability as in a random effects meta-analysis model. By introducing the latent process $\psi(v)$, Yue *et al.* [2012] estimate the probability of a voxel being miscoded. In the HPGRF model, the authors provide correlation estimates between the different emotions. All these quantities can not be obtained by any of the kernel-based methods.

Finally, the Bayesian framework upon which model-based methods are build facilitates the construction of predictive distributions over new studies. This helps producing the so-called *reverse inferences* [Poldrack, 2011], a topic of growing interest in the fMRI community. Traditionally, fMRI studies produce *forward inferences*: for a given task or paradigm, inference is made on the location of the brain response to the task. Reverse inference consists of using the pattern of brain activation to infer which task is most likely to have produced the data. If a neuroscientist has developed a new behavioural experiment (say, on emotion), he/she may indeed want to know whether their task engages the brain's emotion processing system. In such

case the researcher would want an estimate of the probability that the data arose from a population of emotion studies. Kang *et al.* [2014] show that classification based on the HPGRF model outperforms a naive classifier based on MKDA. This result suggests spatial models can captures information in the data that cannot be captured by a MUA.

For all these reasons, we believe that model-based methods have significant merits compared to kernel-based methods. However there are several still open problem for model-based methods and CBMA in general. We discuss those in the following section.

## 3.4 Open problems

There are currently some aspects of CBMA that are being overlooked by both model-based and kernel-based methods. The most important is *publication bias*. Publication bias happens when the studies to which researchers have access are not a representative subset of the total population of studies. For example, one special case of publication bias is the *file drawer* where studies with significant findings are more likely to get published. If present, publication bias can affect the outcome of a meta-analysis and lead to false conclusions. In the field of fMRI there is evidence for the existence of publication biases [David *et al.*, 2013] but there has been no attempt to quantify the effect these biases may have on our meta-analysis estimates. In Chapter 6 we use a simple model to estimate the number of missing studies due to the file drawer in CBMA. Nevertheless, there is still a need to adjust existing methods (or develop new ones) to account for the presence of publication bias.

Another still open problem is meta-regression. Meta-regression extends the simple meta-analysis model to account for study characteristics. When available, such information can improve the fit of a model and give better insights on the data. In CBMA several study characteristics are typically recorded when that data are gathered but it is not yet a commonplace to use this information in meta-analyses. In Chapter 4 we outline a model that uses study characteristics as explanatory variables thus introducing the notion of meta-regression in CBMA.

One interesting problem is how CBMA can be jointly modelled with image data from new fMRI studies. Currently there are no models that connect the image data with the foci and hence it is not possible to use the former as prior information or investigate if the two agree. In Chapter 5 we propose a model for fMRI $T$ statistic data that uses CBMA data as prior information.

Finally, there is only little work done on functional connectivity. Functional

connectivity refers to the dependency between one or more regions of the brain. In CBMA functional connectivity is implied by co-activation, that is, when two regions consistently report activations. In a recent work, Xue *et al.* [2014] use a multivariate Poisson model to induce correlation among the foci in several regions on interest. However, it would be interesting to see if a spatial model can capture these correlations as well.

# Chapter 4

# A Bayesian log-Gaussian Cox process model for CBMA meta-regression

## 4.1 Introduction

In Chapter 3 we have reviewed existing approaches for coordinate-based meta-analysis and have demonstrated the benefits of using model-based methods and in particular approaches that are based on point processes. A limitation, however, of these methods is that they assume that the true intensity function that generated each study is the same for every study. That is, these methods do not allow for systematic differences between studies. Sometimes such differences can be attributed to features that are inherent to the studies, for example the demographic particularities of the samples under investigation. In the present chapter we consider an approach to account for such systematic differences.

The use of study-specific characteristics as explanatory variables in a meta-analysis model is known as meta-regression [Greenland, 1994]. For example, in a meta-analysis of 13 clinical trials on the efficacy of BGC vaccine against tuberculosis, van Houwelingen *et al.* [2002] consider year and altitude as such covariates. Meta-regression is an important facet of meta-analysis, especially when there is appreciable heterogeneity between studies. As noted by Carp [2012], fMRI experiments exhibit incredible variation in the analysis methods applied and hence it is essential to explore the effect that study characteristics have on the analyses outcomes. For example, Carp [2012] found that 59% of a sample of the fMRI literature used a correction for multiple testing. We expect that experiments that fail to adjust for

multiple tests will generally report more foci compared to the ones that adjust, due to the high rate of false positives and thus this is an important factor to account for in a meta-regression.

In this Chapter, we introduce a Bayesian point process model that enables the inclusion of covariates in CBMA thus allowing for meta-regression. The model is outlined Section 4.2 and the algorithm that is used for posterior simulation is described in Section 4.3. In Section 4.4 we test the model on simulated data whereas in Section 4.5 we use it for a meta-analysis of emotion and executive control studies. Finally, we conclude with a discussion of the limitations of this work and future directions.

## 4.2 Model specifications

Suppose that there are a total number of $I$ studies in the meta analysis and that each study comes with a point pattern $\mathbf{x}_i$, a set of foci $x_{ij} \in \mathcal{B} \subset \mathbb{R}^3$, where $\mathcal{B}$ is the support of the analysis, usually set from a standard atlas of the brain, and $j = 1, \ldots, n_i$, where $n_i$ is the number of foci in study $i$. Additionally, suppose that for each point pattern there is a set of $K$ study specific characteristics, $\{z_{ik}\}_{k=1}^{K}$.

We will assume that each point pattern $\mathbf{x}_i$ is the realisation of a Cox point process $\mathbf{X}_i$ defined on $\mathcal{B}$, driven by a random intensity $\lambda_i(\cdot)$. We can then model the intensity function at each point $\xi \in \mathcal{B}$ as:

$$\log \lambda_i(\xi) = \beta_0(\xi) + \sum_{k=1}^{K} \beta_k(\xi) z_{ik}, \tag{4.1}$$

where $\beta_0$ is the baseline intensity parameter, and the $\beta_k$ are the regression coefficients. Equation (4.1) defines a spatial log-linear model over the brain. Foci are more likely to occur in regions of the brain with high intensity values whereas we expect almost no foci in regions as the intensity approaches zero. The exact rates are given by the properties of a Cox process. In particular, given $\lambda_i(\cdot)$, the expected number of foci in any bounded $B \subseteq \mathcal{B}$ is a Poisson random variable with mean $\int_B \lambda_i(\xi) d\xi$ [Møller and Waagepetersen, 2004].

In practice we expect that some covariates will have a global (homogenous) effect. Therefore, we split the covariates into the $K^* \leq K$ that have a local effect and $K - K^*$ that have a global effect:

$$\log \lambda_i(\xi) = \sum_{k=0}^{K^*} \beta_k(\xi) z_{ik} + \sum_{k=K^*+1}^{K} \beta_k z_{ik}, \tag{4.2}$$

where $z_{i0} = 1$ for all $i = 1, \ldots, I$.

A Bayesian model is defined with prior distributions on model parameters, here the functional parameters $\beta_k(\cdot)$, $k = 0, \ldots, K^*$, and scalar parameters $\beta_k$, $k = K^*+1, \ldots, K$. A natural way to proceed is to assume that $\beta_k(\cdot)$ are realisations of Gaussian random fields and the $\beta_k$ have normal distributions. That way, the right hand side of Equation (4.2) is also a Gaussian random field, and each point process is a log-Gaussian Cox process (LGCP) [Møller *et al.*, 1998]. We will assume an isotropic, power exponential correlation structure, that is for points $\xi, \xi' \in \mathcal{B}$ we will have:

$$\mathbb{C}\text{or}\left(\boldsymbol{\beta}(\xi), \boldsymbol{\beta}(\xi')\right) = \exp\left\{-\rho_k ||\xi, \xi'||^{\delta_k}\right\}, \tag{4.3}$$

for $k = 1, \ldots, K^*$, where $\rho_k > 0$ are the correlation decay parameters and $\delta_k > 0$ are the smoothness parameters. The same correlation structure was used by Møller *et al.* [1998] and Møller and Waagepetersen [2003] in the context of LGCPs. Of course, one could consider alternative correlation structures such as the *Matérn* covariance function (see for example Rasmussen and Williams [2005]).

The log-Gaussian Cox process is a flexible model for spatial point data that can account for aggregation [Møller *et al.*, 1998; Møller and Waagepetersen, 2007] or even repulsion between points [Illian *et al.*, 2012a] and has therefore found applications in several fields such as disease mapping [Benes *et al.*, 2002; Liang *et al.*, 2009] and ecology [Møller and Waagepetersen, 2003; Illian *et al.*, 2012b].

By the definition of a Cox process, conditional on $\lambda_i(\cdot)$, $\mathbf{X}_i$ is a Poisson point process on $\mathcal{B}$ [Møller and Waagepetersen, 2004]. The density (Radon-Nikodym derivative) of this point process with respect to the unit rate Poisson process is:

$$\pi\left(\mathbf{x}_i \mid \lambda_i\right) = \exp\left\{|\mathcal{B}| - \int_{\mathcal{B}} \lambda_i(\xi) d\xi\right\} \prod_{x_{ij} \in \mathbf{x}_i} \lambda_i(x_{ij}), \tag{4.4}$$

for $i = 1, \ldots, I$, with $|\mathcal{B}|$ denoting the volume of the brain. We can view $\pi\left(\mathbf{x}_i \mid \lambda_i\right)$ as the density of the sampling distribution of the data; if we further assume independent studies, then we obtain the likelihood for the model as:

$$\mathcal{L}\left(\left\{\beta_k(\cdot)\right\}_{k=0}^{K^*}, \left\{\beta_k\right\}_{k=K^*+1}^{K}\right) = \prod_{i=1}^{I} \pi\left(\mathbf{x}_i \mid \lambda_i\right), \tag{4.5}$$

where $\lambda_i$ is as defined in Equation (4.2). Inference can be then achieved through the

posterior distribution of the model which is given, up to a normalising constant, by:

$$\pi\left(\{\beta_k(\cdot)\}_{k=0}^{K^*}, \{\beta_k\}_{k=K^*+1}^{K} \mid \{\mathbf{x}_i\}_{i=1}^{I}\right) \propto \mathcal{L}\left(\{\beta_k(\cdot)\}_{k=0}^{K^*}, \{\beta_k\}_{k=K^*+1}^{K}\right) \prod_{i=1}^{K^*} \pi(\beta_k(\cdot)) \prod_{k=K^*+1}^{K} \pi(\beta_k), \quad (4.6)$$

where $\pi(\beta_k(\cdot))$ and $\pi(\beta_k)$ are the priors on the functional and scalar parameters, respectively.

### 4.2.1 Posterior approximation

Calculation of the posterior in Equation (4.6) requires the evaluation of the infinite dimensional Gaussian random fields $\beta_k(\cdot)$, $k = 0, \ldots, K^*$, which we approximate with a finite dimensional distribution. Following Møller *et al.* [1998] and Benes *et al.* [2002], we consider the discretisation of the 3D volume with a regular rectangular grid $W \supset \mathcal{B}$. We use $V$ cubic cells (i.e. voxels) in $W$ with volume $A = a^3$, where $a$ is the length of the side. In neuroimaging, analysis with $2\text{mm}^3$ cubic voxels is typical, leading to a grid about 1 million voxels, of which about 200,000 are in the brain or cerebellum. Voxels are indexed $v = 1, \ldots, V$, and the coordinate of $v$ is the location of the center $\boldsymbol{\nu}_v \in \mathbb{R}^3$.

For any $k = 0, \ldots, K^*$, the Gaussian random field $\beta_k(\cdot)$ can be now approximated with a step function which is constant within each voxel $v$ and equal to the value of $\beta_k(\cdot)$ at the location of the center, i.e. $\beta_k(\boldsymbol{\nu}_v)$. Waagepetersen [2004] shows that the accuracy of this approximation improves as $a$ goes to zero. By definition, $\boldsymbol{\beta}_k = [\beta_k(\boldsymbol{\nu}_1), \ldots, \beta_k(\boldsymbol{\nu}_V)]$ are multivariate Gaussian vectors. We parametrise $\boldsymbol{\beta}_k$ as:

$$\boldsymbol{\beta}_k = \mu_k + \sigma_k \mathbf{R}_k^{1/2} \boldsymbol{\gamma}_k, \quad (4.7)$$

where $\mu_k$ are the overall (scalar) means, $\sigma_k$ are the marginal standard deviations, $\mathbf{R}_k$ are the $V \times V$ correlation matrices with elements $(\mathbf{R}_k)_{ij} = \exp\left\{-\rho_k ||\boldsymbol{\nu}_i, \boldsymbol{\nu}_j||^{\delta_k}\right\}$, and $\boldsymbol{\gamma}_k$ are the *a priori* $\mathcal{N}_V(\mathbf{0}, \mathbf{I}_V)$ vectors, $k = 0, \ldots, K^*$. The same parametrisation is used by Møller *et al.* [1998], Christensen and Waagepetersen [2002] and is advocated by Christensen *et al.* [2006] because it allows for computationally efficient posterior simulations. For the purposes of this work we will only consider the case where $\delta_k = 2$ i.e. the *Gaussian correlation* function. This choice may seem rather simplistic but is justified by the sparsity of points in CBMA data and ubiquitous use of Gaussian smoothing kernels in neuroimaging data analysis.

Priors for the $V$-vectors $\boldsymbol{\gamma}_k$ are induced by the parametrisation of Equation

(4.7). For the remaining parameters we will assume weekly informative priors: $\mu_k \sim \mathcal{N}\left(0, 10^8\right)$, $\sigma_k \sim \mathcal{N}\left(0, 10^8\right)$, $\beta_k \sim \mathcal{N}\left(0, 10^8\right)$ and $\rho_k \sim \text{Uni}\left[0, 100\right]$[1].

Once the latent fields are approximated, one can also approximate $\lambda_i$ with a step function as before. The intensities at the center of each voxel are given by:

$$\boldsymbol{\lambda}_i = \exp\left\{\sum_{k=0}^{K^*}\left(\mu_k + \sigma_k \mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right) z_{ik} + \sum_{i=K^*+1}^{K} \beta_k z_{ik}\right\}, \tag{4.8}$$

where $\boldsymbol{\lambda}_i$ is the $V$-vector, the discretised intensity. We will write $\lambda_{iv} = (\boldsymbol{\lambda}_i)_v$ for the $v$-element of study $i$'s intensity, and note we require $z_{i0} = 1$ to capture the mean effect. The approximated posterior is:

$$\pi\left(\boldsymbol{\theta} \mid \{\mathbf{x}_i\}_{i=1}^{I}\right) \propto \prod_{i=1}^{I}\left[\exp\left\{-\sum_v A_v \lambda_{iv}\right\}\prod_{j=1}^{n_i}\lambda_{iv(x_{ij})}\right]\pi(\boldsymbol{\theta}), \tag{4.9}$$

where $\boldsymbol{\theta} = \left\{\{\mu_k\}_{k=1}^{K^*}, \{\sigma_k\}_{k=1}^{K^*}, \{\rho_k\}_{k=1}^{K^*}, \{\boldsymbol{\gamma}_k\}_{k=1}^{K^*}, \{\beta_k\}_{k=K^*+1}^{K}\right\}$, $A_v$ takes on the value $A$ when $\boldsymbol{\nu}_v \in \mathcal{B}$ and 0 otherwise, $v(x_{ij})$ is the index of the voxel containing $x_{ij}$, and $\pi(\boldsymbol{\theta})$ is the joint prior distribution of the parameters. The posterior distribution in Equation (4.9) is still analytically intractable due to the presence of an unknown normalising constant and thus we need to resort to Monte Carlo simulation or approximation techniques to obtain samples from it.

## 4.3 Sampling algorithm details

Bayesian methodology for inference on LGCPs can be broadly divided into two main categories: simulation based approximations of the posterior such as Markov Chain Monte Carlo [Møller *et al.*, 1998] and *elliptical slice sampling* [Murray *et al.*, 2010], and deterministic approximations to the posterior such as *integrated nested Laplace approximations* [Illian *et al.*, 2012a; Simpson *et al.*, 2016, INLA] and *variational Bayes* [Jaakkola and Jordan, 2000]. In a recent study, Taylor and Diggle [2014] compare the Metropolis-adjusted Langevin (MALA) algorithm with INLA and find that both methods give similar results. In our application, we choose to use simulation based methods because application on a 3D problem is more straightforward.

Of course, building an algorithm for such high-dimensional problem is non-trivial. Girolami and Calderhead [2011] showed that of all possible strategies, their Riemann Manifold Hamiltonian Monte Carlo (RMHMC) sampler is the computa-

[1]This prior may not be uninformative for applications with a different scale of distances

tionally most efficient for LGCPs in a 2D setting. Unfortunately, application in this problem (3D setting) is prohibitive as it would require the inversion of a huge $V \times V$ matrix. Therefore, we choose to use the standard Hamiltonian Monte Carlo [Duane *et al.*, 1987; Neal, 2011, HMC] algorithm which Girolami and Calderhead [2011] found to be the second most efficient.

HMC initially appeared in the physics literature by Duane *et al.* [1987] under the name *Hybrid Monte Carlo*, and later emerged into statistics literature by Neal [2011]. We now give a brief description of the method. HMC emulates the evolution of a particle system which is characterised by its position ($\mathbf{q}$) and momentum ($\mathbf{p}$) over time. In our case, $\mathbf{q}$ will be the parameter vector of interest $\boldsymbol{\theta}$ and $\mathbf{p}$ will be introduced artificially from a $\mathcal{N}_d(0, \mathbf{M})$ distribution, with $d$ being the dimensionality of the problem and $\mathbf{M}$ the mass matrix. The dynamics of the system are described by a set of differential equations, known as Hamilton's equations.

HMC alternates between moves for the position vector $\boldsymbol{\theta}$ and the momentum vector $\mathbf{p}$ based on Hamilton's equations. If the solutions of the equations can be found analytically then moves will be deterministic; if not, numerical integration is required and an acceptance/rejection step must be performed to account for integration error. Integration is done in fictitious time $\epsilon L$, where $\epsilon$ is the *stepsize* and $L$ is the *number of steps*. Typically the *leapfrog integrator* is employed, which for $L = 1$ and starting form time $t$ is performed as [Neal, 2011]:

$$\begin{aligned} \mathbf{p}\left(t + \frac{\epsilon}{2}\right) &= \mathbf{p}\left(t\right) + \frac{\epsilon}{2}\nabla_{\boldsymbol{\theta}} \log \pi\left(\boldsymbol{\theta}(t) \mid \{\mathbf{x}_i\}_{i=1}^{I}\right) \\ \boldsymbol{\theta}\left(t + \epsilon\right) &= \boldsymbol{\theta}\left(t\right) + \epsilon \mathbf{M}^{-1}\mathbf{p}\left(t + \frac{\epsilon}{2}\right) \\ \mathbf{p}\left(t + \epsilon\right) &= \mathbf{p}\left(t + \frac{\epsilon}{2}\right) + \frac{\epsilon}{2}\nabla_{\boldsymbol{\theta}} \log \pi\left(\boldsymbol{\theta}(t+\epsilon) \mid \{\mathbf{x}_i\}_{i=1}^{I}\right). \end{aligned} \tag{4.10}$$

Overall, if the method is applied correctly, it will produce a time-reversible Markov chain that has the desired distribution $\pi\left(\boldsymbol{\theta} \mid \{\mathbf{x}_i\}_{i=1}^{I}\right)$ as its stationary distribution. As we show in Appendix A.1, gradient expressions are available in closed form for all model parameters including correlation parameters $\rho_k$. We therefore choose to update all $\{\mu_k, \sigma_k, \rho_k, \boldsymbol{\gamma}_k\}_{k=0}^{K^*}$ and $\{\beta_k\}_{k=K^*+1}^{K}$ jointly with HMC. The solutions to Hamilton's equations are not available analytically so we need to use the Leapfrog integrator and include an accept/reject step at the end of it.

The procedure requires the specification of a stepsize $\epsilon$ and a total number of leapfrog steps $L$. Hoffman and Gelman [2014] show how tuning can be achieved automatically but when we applied this method to our problem running time was increased substantially. Therefore we use an alternative approach to tune these parameters. The stepsize is automatically adjusted during the burn-in phase of the

HMC to give an overall acceptance rate close to the 65% suggested by Neal [2011]. In particular, if $\epsilon_t$ is the stepsize at iteration $t$ and $q_{t_1}$ is the acceptance rate over the past $t_1$ iterations, then every $t_2$ iterations we calculate the new stepsize $\epsilon'_t$ as:

$$\epsilon'_t = \begin{cases} 0.9\epsilon_t & q_{t_1} < 0.60 \\ \epsilon_t & 0.60 \leq q_{t_1} \leq 0.70 \\ 1.1\epsilon_t & q_{t_1} > 0.70 \end{cases} . \tag{4.11}$$

Specifically we use $t_1 = 100$ and $t_2 = 10$ A similar approach is empoyed by Marshall and Roberts [2012] for MALA. The latter (number of leapfrog steps), is always fixed to $L = 50$. We took this approach because we found that, for our LGCP application, the mixing properties of the algorithm scale linearly with $L$ but also with the total number of HMC iterations. Hence one can use a relatively large $L$ and few iterations or relatively smaller $L$ and more iterations, the total computation time staying relatively constant.

The last tuning parameter in the HMC algorithm is the variance-covariance matrix of the zero mean normal momentum parameters, $\mathbf{M}$. To our knowledge, there is only limited off the shelf methodology as to how $\mathbf{M}$ can be adjusted and so a good starting place is to set $\mathbf{M} = \mathbf{I}$. Neal [1996] suggests that if an estimate of the posterior variance $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ is available then a good practice is to set $\mathbf{M} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{-1}$. In principle, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}$ can be estimated during the burn-in phase of HMC but in practice this is not possible due to the dimensionality of the problem. In our simulations, we found that the mean posterior variance of the elements of the $\boldsymbol{\gamma}_k$ was higher compared to the scalar parameters, followed by $\beta_k$ or $\sigma_k$ and then $\rho_k$. Especially for the $\rho_k$ the scale is typically much smaller compared to the other parameters in our applications and so we use $100 \times \rho_k$ instead of $\rho_k$. After the reparametrisation we found that setting the mass for parameters of $\boldsymbol{\gamma}_k$, $\beta_k$, $\sigma_k$ and $\rho_k$ equal to 1, 3, 3 and 10 respectively worked well in all implementations.

The most computationally demanding part of the algorithm is the the calculation of the large matrix-vector products $\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k$ appearing in the intensity functions of Equation (4.8). Luckily, an elegant solution to this problem is given by Møller *et al.* [1998] based on *circulant embedding* that was first proposed by Dietrich and Newsam [1993] and Wood and Chan [1994]. The key to the approach is the linear algebra result that a circulant matrix has the discrete Fourier basis as its eigenvectors. $\mathbf{R_k}$ is not circulant but is block toeplitz and can be embeded in a $(2V) \times (2V)$ matrix that is circulant. Thus the matrix square root, inversion and multiplication can be accelerated by using (the highly efficient) discrete Fourier

transform (DFT) of the embedded matrix and manipulating Fourier coefficients, followed by inverse DFT and extracting the appropriate sub-matrix/sub-vector. See Rue and Held [2005] for more details.

We close this section by stressing that despite the massive dimensionality of the parameter vector, the problem has a very high degree of parallelisation. Intensities can be evaluated in blocks of thousands of voxels simultaneously making the algorithm suitable for implementation in a graphical processing unit (GPU). The most computationally intensive part of our model, namely operations with DFTs, is also amenable to parallelisation and there already exist libraries such as NVIDIA's cuFFT library that are designed for this specific task. Overall, we believe that implementation of the log-Gaussian Cox process model described above will soon become a routine task for any moderately powerful GPU device.

## 4.4 Simulation studies

We consider two simulation setups. In the first we draw samples directly from the log-Gaussian Cox process whereas in the second we create synthetic studies based on a different model to assess the robustness of our method to model misspecification. For consistency, all processes are defined on the same brain atlas used in the application of Section 4.5, consisting of $216,040$ $2\text{mm}^3$ cubic voxels. The average number of foci per simulated study is kept low (mean number of foci per study is 5) to resemble the sparsity of points observed in real CBMA data. Finally, the total number of studies is fixed to 200 in both analyses, similar to the sample sizes available in real applications [Kang *et al.*, 2011, for example].

### 4.4.1 Setup 1

In this setting we simulate 200 studies, with two spatially varying covariates that account for the mean of two groups of studies, and two non-spatially varying covariates. For $i = 1, \ldots, 200$ we set:

$$\lambda_{iv} = \exp\left\{\sum_{k=1}^{2}\left(\mu_k + \sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_v\right)z_{ik} + \sum_{i=3}^{4}\beta_k z_{ik}\right\}, \tag{4.12}$$

where $z_{i1} \sim \text{Bernoulli}(0.5)$, $z_{i2} = 1 - z_{i1}$, $z_{i3} \sim \text{Uniform}[-1, 1]$ and $z_{i4} \sim \text{Bernoulli}(0.5)$. Note that this parametrisation of the covariates implies existence of two types of studies, say type 1 and 2, with different spatially varying means and the effect of one continuous and one categorical covariate. The expected total number of foci is

3.99 and 4.16 for studies of type 1 and 2 respectively. We draw $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$ from their $\mathcal{N}_V(0, \mathbf{I})$ prior and fix the values of the scalar parameters shown in Table 4.1. We run the HMC algorithm of Section 4.3 for 10,000 iterations, discarding the first 4,000 as a burn-in and save every 6 iterations for a total of 1,000 saved posterior samples. This took roughly 14 hours on an NVIDIA Tesla K20c GPU card.

Results are summarised in Table 4.1 and Figures 4.1 and 4.2. In Table 4.1 we see that the scalar parameters are estimated accurately despite the sparsity of points in the realisations. The 95% credible intervals contain the true values of all the parameters in the setup. Some traceplots for these parameters can be found in Section A.2 of the Appendix.

For $z_{i3} = z_{i4} = 0$, the median expected number of points is 3.97 (95% CI [3.84,4.10])for type 1 and 4.61 (95% CI [4.46,4.78]) for type 2. These values are very similar to the values we observe in the simulated dataset, that is 3.98 for type 1 and 4.53 for type 2. This indicates that our model does a good job fitting the data. The shape of the latent Gaussian fields $\mu_k + \sigma_k \mathbf{R}_k^{1/2} \boldsymbol{\gamma}_k$ is generally captured for both types as can be seen in Figures 4.1 and 4.2. In particular, we can see that the maxima in the true and estimated images appear roughly in the same locations. The same cannot be said about the other values but this is expected given the dearth of information in regions of low intensity.

Table 4.1: Posterior summaries of the scalar parameters of the LGCP model, fit to the simulated data of Section 4.4.1. Results are based on 1,000 posterior draws. The values for the correlation parameters $\rho_1$, $\rho_2$ are multiplied by 100. The values for $\beta_1$ and $\beta_2$ are multiplied by 10.

| **Parameter** | **True Value** | **Posterior median** | 95% **credible interval** |
|---|---|---|---|
| $\mu_1$ | -13.7 | -13.72 | [ -13.99 , -13.48 ] |
| $\mu_2$ | -14.2 | -14.14 | [ -14.47 , -13.86 ] |
| $\sigma_1$ | 1.2 | 1.19 | [ 1.01 , 1.38 ] |
| $\sigma_2$ | 1.6 | 1.61 | [ 1.43 , 1.81 ] |
| $\rho_1$ | 1 | 0.93 | [ 0.69 , 1.27 ] |
| $\rho_2$ | 2 | 2.30 | [ 1.69 , 3.15 ] |
| $\beta_3$ | 2 | 1.44 | [ 0.22 , 2.52 ] |
| $\beta_4$ | 1 | 0.95 | [ 0.32 , 1.65 ] |

### 4.4.2 Setup 2

In this setup we create datasets with a pattern of points that follows brain structures of interest. Again there are two types of studies, say type 1 and type 2. For each

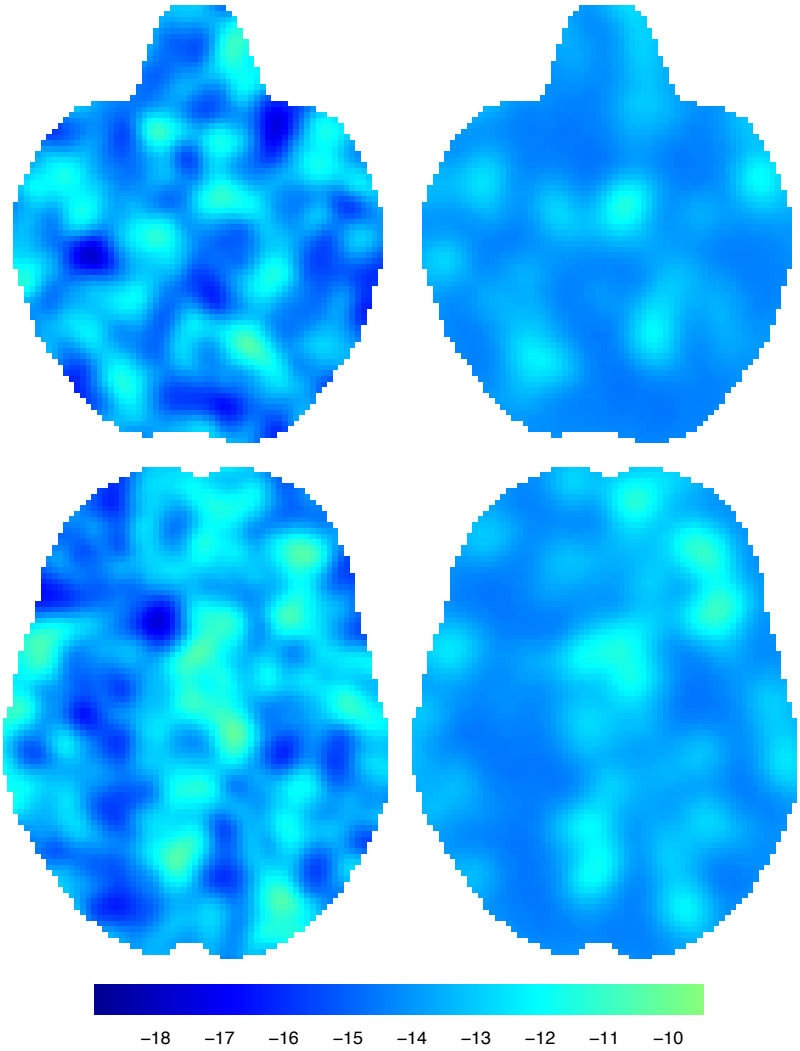



Figure 4.1: Some true (left) and estimated (right) latent Gaussian fields for type 1 in the simulation setup 1 of Section 4.4.1. Rows 1 and 2 correspond to axial slices $z = -22$ and $z = 4$ respectively. The most intense regions show similar structure, while less intense regions have too few points to learn the intensity.

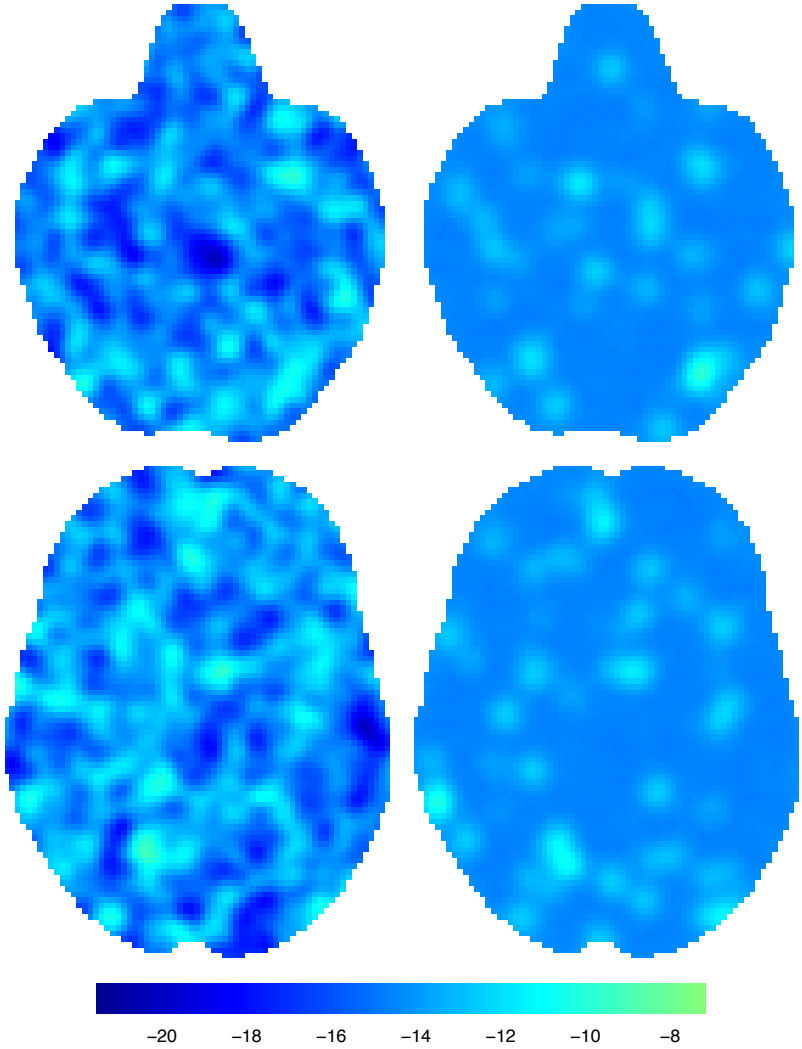



Figure 4.2: True (left) and estimated (right) latent Gaussian fields for type 2 in the simulation setup 1 of Section 4.4.1. Rows 1 and 2 correspond to axial slices $z = -22$ and $z = 4$ respectively. The most intense regions show similar structure, while less intense regions have too few points to learn the intensity.

study $i$, $i = 1, \ldots, 200$, we generate the total number of points from a Negative Binomial distribution with mean $\mu = 6 + 2z_{i3} - \mathbf{1}_{\{z_{i4}=0\}} + \mathbf{1}_{\{z_{i4}=1\}}$ and variance $\mu^2/20$. For the covariates, $z_{i3} \sim \text{Uni}[-1, 1]$ and $z_{i4} \sim \text{Bernoulli}(0.5)$. Once we know the exact number of foci per study, we assign the study uniformly at random to one of the 2 types and the distribute its foci as follows. For type 1, foci appear systematically in the following regions: each focus can be observed in the right amygdala ($B_R$) with probability 55%, the orbifrontal cortex ($B_C$) with probability 30% or anywhere else in the brain with probability 15%. The configuration for type 2 differs in that most of the points will go to the left amygdala ($B_L$) instead of the right amygdala. If a focus is assigned to one of the three broad regions, the exact location has a uniform distribution over the region. In the bottom left panel of Figures 4.3 and 4.4 the regions in red and blue correspond to the left and right amygdala respectively while the orbifrontal cortex coloured in green.

HMC is run for 10,000 iterations, discarding the 4,000 first as a burn-in and saving every 6 to obtain a total of 1,000 samples from the posterior. The run took 15 hours on a Tesla K20c GPU card.

Results are shown in Figures 4.3 and 4.4 where in the top rows we see median posterior log-intensities for the two types, in different axial slices. In both cases, we find that the regions with the highest intensities are the amygdalae and that the orbifrontal cortex is a region of high posterior intensity as well. The median expected number of points is 5.81 for type 1 (95% CI [5.36,6,32]) and 6.45 for type 2 (95% CI [5.97,6.97]). The observed values are 6.27 and 6.73 respectively.

Conditional on there being exactly one focus, we can estimate the probability that this focus appears in any subset $B \subseteq \mathcal{B}$ as $\int_B \lambda(\xi)\, d\xi / \int_{\mathcal{B}} \lambda(\xi)\, d\xi$. Using the posterior draws obtained from the HMC algorithm, we can obtain the posterior distribution of any such quantity. For our simulated type 1 data we find that the median posterior probability of observing a focus in the right amygdala ($B_R$) is 0.43 (95% CI [0.40,0.48]). For type 2, the probability of observing a focus in the left amygdala ($B_L$) is 0.42 (95% CI [0.39,0.46]). For the orbifrontal cortex ($B \equiv B_C$) the median posterior probabilities are 0.25 for type 1 and 0.23 for type 2, with 95% credible intervals $[0.22, 0.28]$ and $[0.20, 0.26]$ respectively. We therefore see that the model underestimates the probabilities for $B_R$, $B_L$ and $B_C$. This bias can be attributed to the smoothness that is imposed by our parameter $\delta$ thus leading to increases intensities just outside these regions as well as regions where noise foci appear.

An interesting question one may ask is which are the regions of the brain that are activated by one type or the other, but not both. To answer this, one can

construct the mean standardised posterior difference map computed as the ratio of the posterior mean of the difference $(\boldsymbol{\beta}_1)_v - (\boldsymbol{\beta}_2)_v$, to the posterior standard deviation of that difference: $\frac{(\boldsymbol{\beta}_1)_v - (\boldsymbol{\beta}_2)_v}{\text{sd}\left((\boldsymbol{\beta}_1)_v - (\boldsymbol{\beta}_2)_v\right)}$. Extreme negative or positive values are evidence of differences between the two types. We show the difference map in the bottom right panels of Figures 4.3 and 4.4. As we see, the model distinguishes the the two types in the amygdala but the differences are small in the rest of the brain. This is a very interesting feature of the model, especially for applications in CBMA where researchers are sometimes interested in comparing a similar process in a different domain, see for example Rottschy *et al.* [2012] or Section 4.5 for an application. An alternative way to do the comparison would be to use the posterior intensity draws to find $\mathbb{P}\left(\boldsymbol{\lambda}_v^1 - \boldsymbol{\lambda}_v^2 > \epsilon\right) \approx \frac{\sum_{n=1}^{N} \mathbf{1}_{\{\lambda_v^{1n} - \lambda_v^{2n} > \epsilon\}}}{N}$, where $\epsilon$ is a threshold difference, $\lambda_v^{kn}$ is the posterior intensity for type $k$ in voxel $v$ obtained from the $n$-th iteration of the algorithm, and $N$ is the total number of HMC draws. However, in this approach it is hard to choose the threshold difference $\epsilon$.

## 4.5 Application: meta-analysis of emotion and executive control studies

In this Section we apply our model to a real meta-analysis data set.

### 4.5.1 Data despcription

Our dataset consists of 1,193 neuroimaging studies. The studies were conducted between 1985 and 2015 and the average number of participants was 16. Of these, 855 studies are on emotion and the remaining 338 are on executive control. The sample has a total of 10,266 foci, 6,112 from emotion (7.15 on average) and 4,154 from executive control (12.289 on average). Figure 4.5 is a graphical representation of the data. We see that even though there is some clustering, the foci are distributed throughout the brain. Our application will focus on identifying the regions of consistent activation across studies and infer on possible difference between the two types.

### 4.5.2 Algorithm details and convergence diagnostics

We use the same parametrisation as in Section 4.4: we have 2 spatially varying intercepts, one for emotion and one for executive control. We run the HMC for 15,000 iterations in total, discarding the first 5,000 as a burn-in. The total number of leapfrog steps is set to $L = 50$ and the stepsize is initialised at $\epsilon = 0.0001$. We

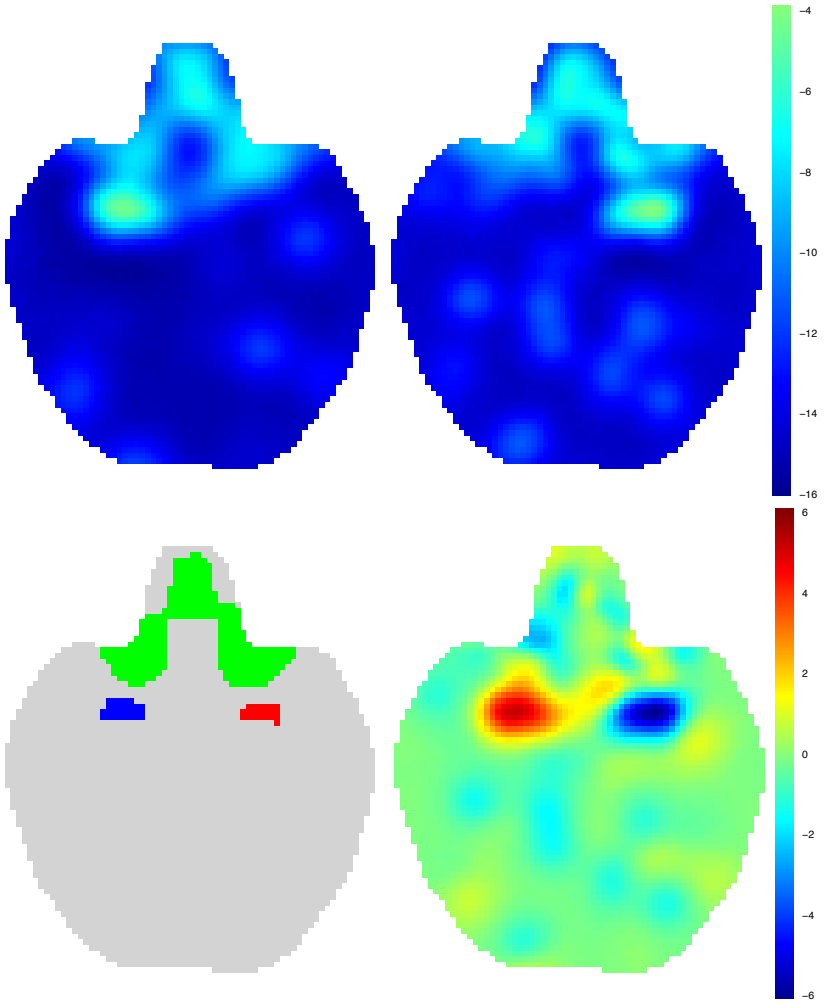



Figure 4.3: Results for simulation setup 2 of Section 4.4.2. Top row are the estimated log-intensities for type 1 and type 2 respectively. The right panel of the bottom row is the standardised mean posterior difference between the two latent Gaussian fields; bright colours indicate areas mostly activated by type 1 process. The left panel of the bottom row shows the regions of the brain systematically activated by the two processes; red for type 1, blue for type 2 and green for both. All images correspond to slice $z = -24$ of the brain.

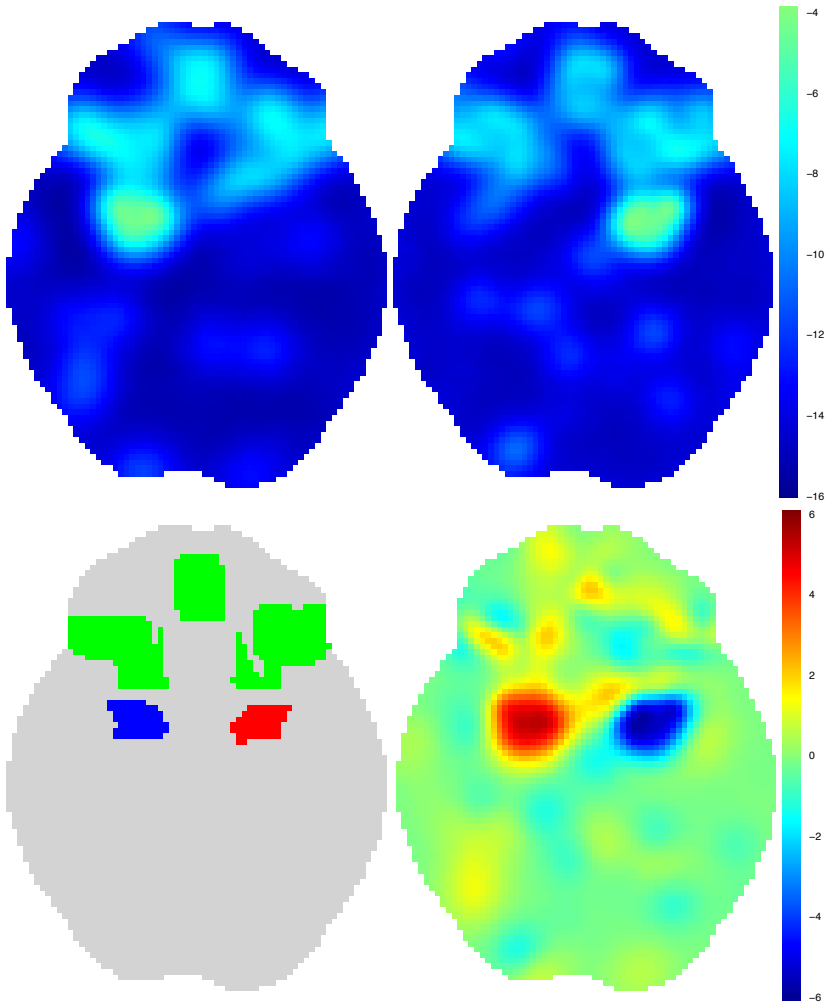



Figure 4.4: Results for simulation setup 2 of Section 4.4.2. Same layout as Figure 4.3 (see that caption for details), except here slice $z = -16$ of the brain is shown.

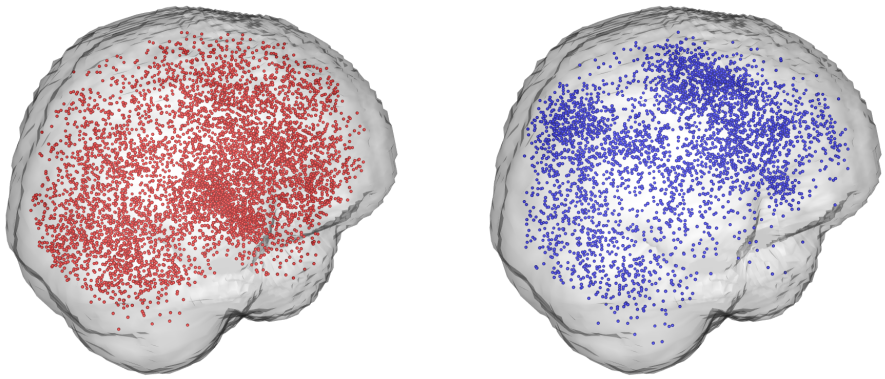


Figure 4.5: Graphical representation of the meta-analysis dataset. Data consist of foci from 1,193 studies of two types, namely 855 emotion (left,red) and 338 exectutive control (right, blue) studies. The overall number of foci is 6,112 and 4,154 for emotion and executive control respectively, for a combined total of 10,266 foci. The code used to generate the figure is courtesy of Jian Kang.

use a diagonal mass matrix with units specified as explained in Section 4.3. For the long vectors $\boldsymbol{\gamma}_k$, $k = 1, 2$ we set the variance equal to 1. For the scalar parameters $\sigma_k$, $\mu_k$ and $\rho_k$ we use mass parameters equal to 3, 3 and 10 respectively.

Convergence of the HMC chain is assessed visually by inspection of posterior traceplots for the model parameters. We run a total of 3 HMC chains in order to examine if they all converge to the same values. Posterior traceplots, along with autocorrelation plots are shown in Section A.3 of the Appendix. Due to the large number of parameters we only focus on the scalars $\mu_k$, $\sigma_k$ and $\rho_k$, as well as intensities in voxels $v_1$ and $v_2$ where the highest median posterior values are observed for the two types. Integrated intensities over the entire brain are also examined. Results indicate that our chains have converged to their stationary distribution. This is verified by the fact that posterior values from the 3 different runs overlap one with another for all the quantities the we plot. Finally it is worth noting that our chains show very good mixing properties since autocorrelation falls to low values after only a small number of iterations.

### 4.5.3 Results

In Figure 4.6 we plot median posterior intensities for emotion in several axial slices of the brain. Our results look qualitatively similar to the results obtained by Kober *et al.* [2008], Yue *et al.* [2012] and Kang *et al.* [2011] in their meta-analyses of

emotion studies. The expected number of foci is 7.15 (95% CI [6.99,7.33]) and the regions mostly activated by emotions are the right and the left amygdala where the peak intensities in Figure 4.6 appear. Executive control processing generally recruits more regions of the brain and hence the median expected number of foci is 12.30 (95% CI [11.92,12.68]). The main effects are localised in the right and left cerebral cortex as can be seen in Figure 4.7.

Several quantities of interest can be obtained from our model, based on the properties of the spatial Poisson process. For example, one may calculate the probability of observing at least one focus in any voxels, region of interest (ROI) or the entire brain. In Figure 4.9 we show the posterior distribution of $\mathbb{P}\left(N_{\mathbf{X}}\left(B\right) \geq 1\right)$, the probability of observing at least one focus in B, for several ROIs $B$. The division of the brain in ROIs is done according to the Harvard-Oxford atlas [Desikan *et al.*, 2006]. A full brain analysis can be found in Section A.4 of the Appendix. Note that this type of inference cannot be easily obtained from kernel-based methods such as MKDA [Wager *et al.*, 2007] or ALE [Eickhoff *et al.*, 2012] and therefore is a relative merit of our point process model.

Comparison of the two types can made as described in Section 4.4.2. The mean standardised posterior difference of the two intercepts is show in Figure 4.8. Generally, we see that the two types have entirely distinct localisations and the main differences are observed where the main effects appear that is, the amygdlala and the cerebral cortex.

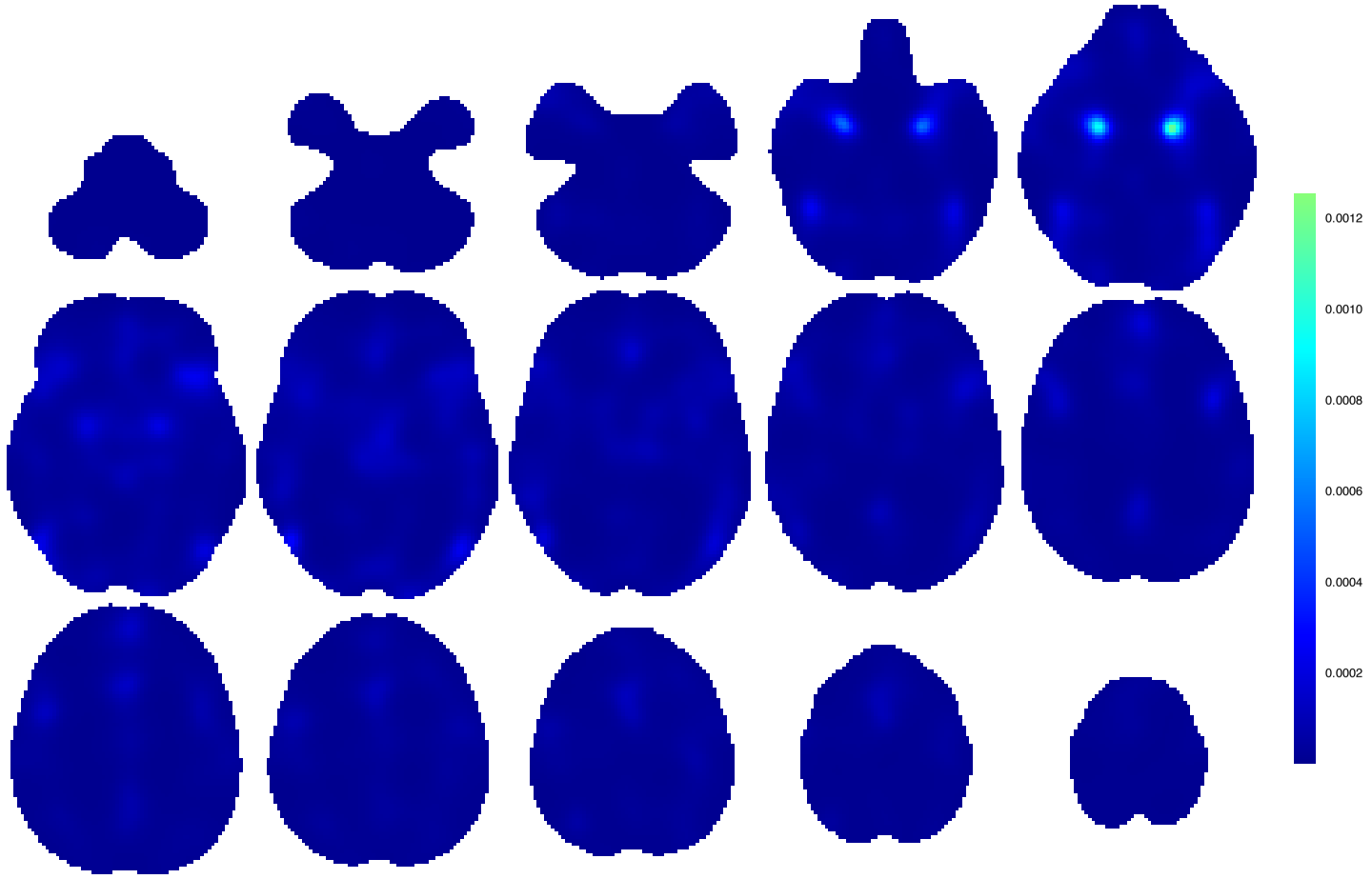



Figure 4.6: Emotion results. The figure presents the median posterior intensity at each voxel for several axial slices, as obtained from 1,000 independent draws from the posterior. Top row (left to right): $z = -50$, $z = -42$, $z = -32$, $z = -24$ and $z = -16$. Middle row (left to right): $z = -8$, $z = +2$, $z = +10$, $z = +18$ and $z = +28$. Bottom row (left to right): $z = +36$, $z = +44$, $z = +52$, $z = +62$ and $z = +70$.

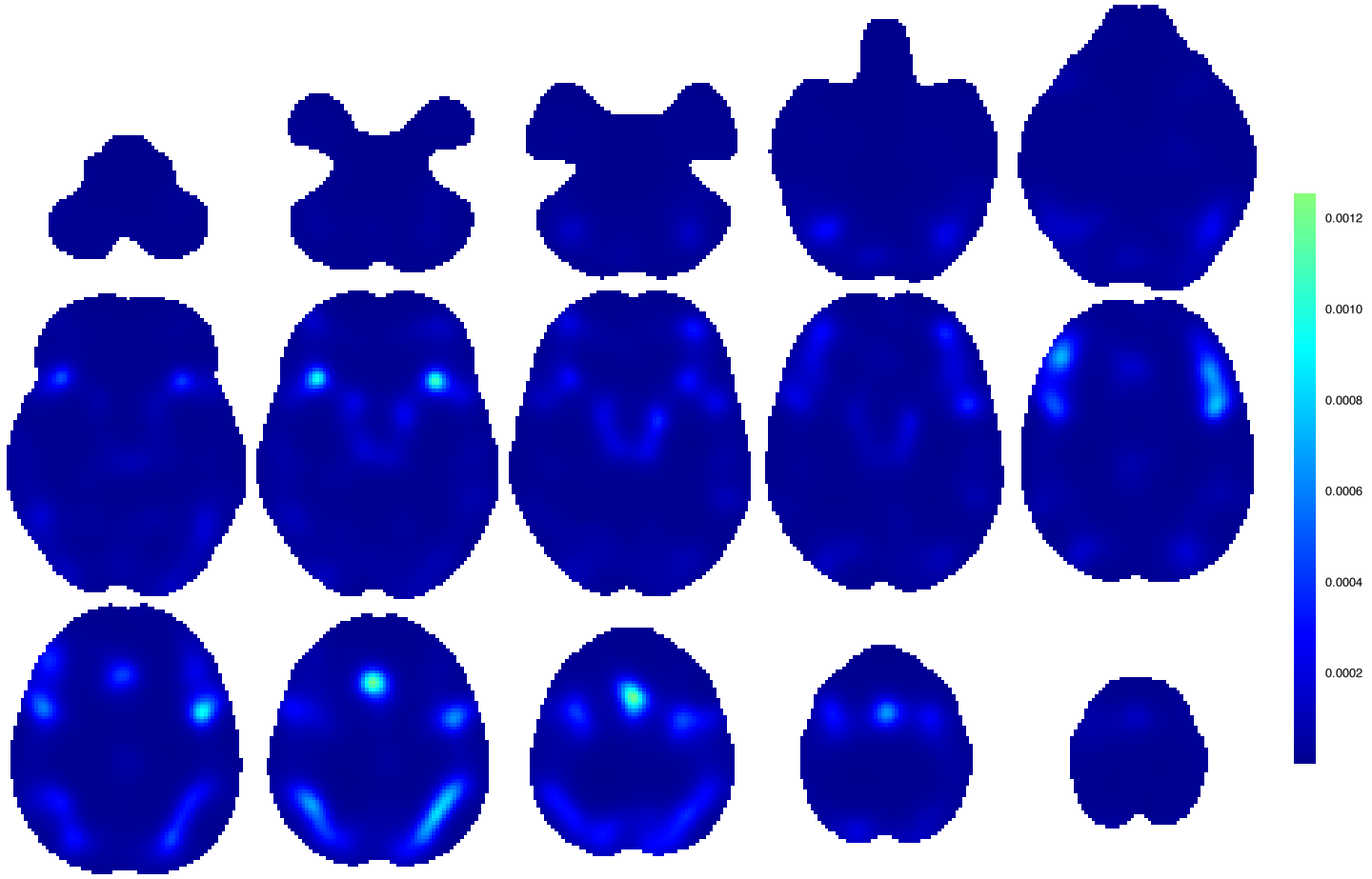



Figure 4.7: Executive control results. The figure presents the median posterior intensity at each voxel for several axial slices, as obtained from 1,000 independent draws from the posterior. Top row (left to right): $z = -50$, $z = -42$, $z = -32$, $z = -24$ and $z = -16$. Middle row (left to right): $z = -8$, $z = +2$, $z = +10$, $z = +18$ and $z = +28$. Bottom row (left to right): $z = +36$, $z = +44$, $z = +52$, $z = +62$ and $z = +70$.

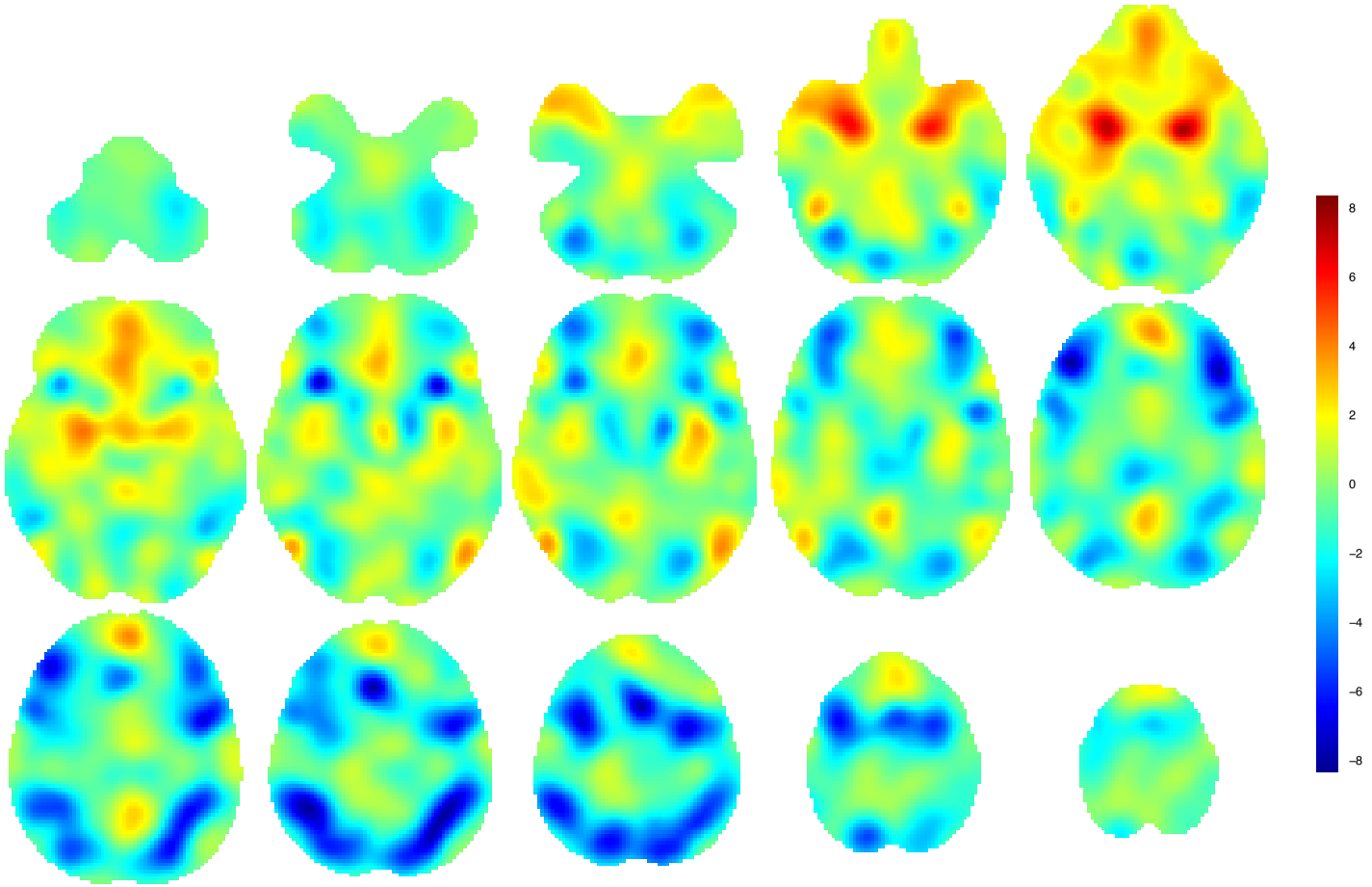



Figure 4.8: Comparison of the two types. The figure presents mean standardised posterior differences of the latent gaussian fields at each voxel for several axial slices, as obtained from 1,000 independent draws from the posterior. Top row (left to right): $z = -50$, $z = -42$, $z = -32$, $z = -24$ and $z = -16$. Middle row (left to right): $z = -8$, $z = +2$, $z = +10$, $z = +18$ and $z = +28$. Bottom row (left to right): $z = +36$, $z = +44$, $z = +52$, $z = +62$ and $z = +70$.

### 4.5.4 Model assessment

Posterior predictive checks [Gelman *et al.*, 1996] can be found in Section A.5 of the Appendix. In particular, we compare first and second order properties of observed data with samples obtained from the posterior predictive distribution. Results indicate some weakness in capturing the second order properties of the data but are overall found satisfactory.

## 4.6 Discussion

In this chapter, we have presented a new coordinate-based meta-analysis model, extension of the log-Gaussian Cox process model of Møller *et al.* [1998]. To our knowledge, this is the first application of the LGCP with several realisations in a 3D problem. Note that even though our application is focused on neuroimaging, the method is directly applicable to any multi-type point pattern problem.

The model has an appealing interpretation being a spatial GLM and several interesting inferences can be obtained based on the properties of the spatial Cox process. A significant advantage of our approach compared to existing methods is the inclusion of covariates in the analysis thus introducing the notion of meta-regression in CBMA. Another very interesting feature of the model is that it allows multi-type comparison directly from the posterior without having to run the model several times.

The main weakness of our approach is the large amount of computational effort required. Nevertheless, the proposed HMC algorithm exhibits quick convergence and good mixing properties and hence less samples from the posterior are required. Additionally, implementation of the method on a GPU vastly reduces computation time and thus makes the method applicable on big meta-analysis problems.

Application of the method on a meta-analysis of emotion and executive control studies has given valuable insights on the data. In particular, we have found that the main structures activated by emotion processing are the right and left amygdala, a finding which is consistent with previous studies on the topic. Executive control functions engage more regions compared to emotion and the main areas activated are the right and left cerebral cortex. Furthermore, comparison of the two types has allowed the detection of several regions in which they differ significantly.

There are several ways in which our model can be extended. One option

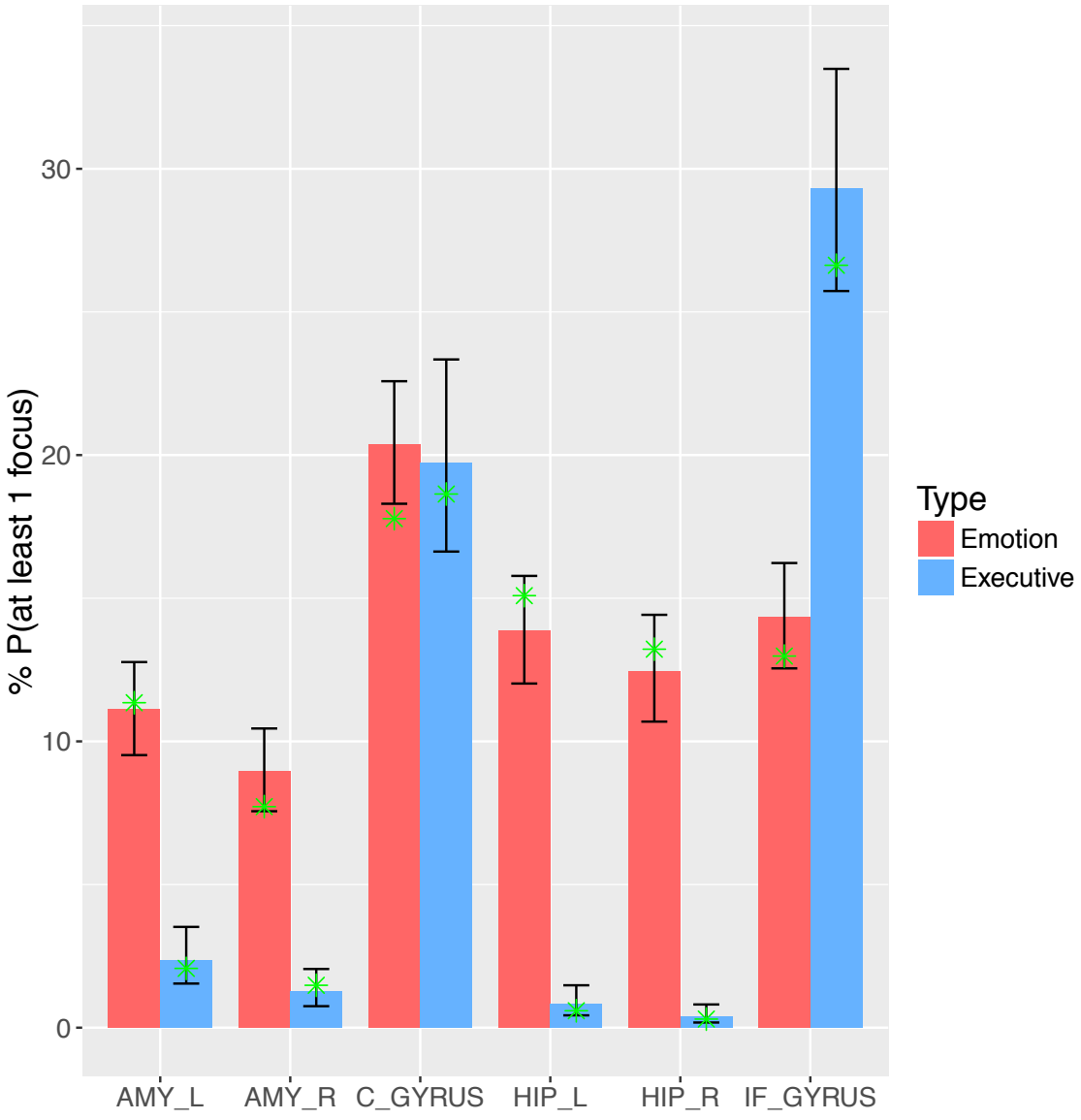



Figure 4.9: Posterior % probabilities of observing at least one focus for several ROIs. The bars represent the median posterior values, red for emotion and blue for executive control. The 95% posterior credible intervals are shown in black. Green asterisks are the empirical values as obtained from the data. The ROIs are (left to right): left amygdala, right amygdala, anterior cingulate gyrus, left hippocampus, right hippocampus and inferior frontal gyrus pars opercularis.

would be consider a random effects LGCP model where:

$$\log \lambda_i(\xi) = \sum_{k=0}^{K^*} \beta_k(\xi) z_{ik} + \sum_{k=K^*+1}^{K} \beta_k z_{ik} + \epsilon_i, \tag{4.13}$$

where $\epsilon_i$ are the random effect terms with $\epsilon_i \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$. This extension would account for further variability among the observed point patterns. Gradient expression are still available in closed form and hence inference for this model is still feasible by jointly updating the random effects parameters $\{\epsilon_i\}_{i=1}^{I}$ along with the remaining parameters of the LGCP.

Another possibility is to use some additional information about the foci such as $p$-values or $T$ scores. These values can be attached as marks to the existing point patterns in order to improve estimation of the intensity function. Such an approach can enrich the inferences obtained from a meta-analysis by characterising the magnitude of activation in each region as opposed to the localisation of activations, which is the question that current methods address. Finally, it would be interesting to see how point pattern data can be combined with IBMA data so that a meta-analysis can take advantage of all the available information on a topic. Radua *et al.* [2012] have already done some work for kernel-based methods but the problem still needs to be tackled for spatial models as well.

# Chapter 5

# A Bayesian spatial model for group fMRI studies

## 5.1 Introduction

The statistical analysis of fMRI data is typically done using a massively univariate approach (MUA). In this approach, a univariate regression model is fit to the time series of the BOLD signal observed at each voxel seperately. The regressors of the model represent the presence of various conditions or tasks that participants were asked to perform during the experiment and therefore at a first level we are able to investigate the presence of task effects for each individual voxel and participant seperately. This is achieved with a $t$ test performed at each voxel using contrasts of the estimated regression coefficients chosen appropriately to represent a given hypothesis related to the tasks [Mumford and Nichols, 2006]. Typically, we compare the BOLD responses during a task relative to the baseline condition but it is possible to compare two different tasks (e.g. fear provoked by visual stimuli as opposed to fear provoked by acoustic stimuli). For each participant, the resulting $T$ image of the values of the $t$ tests at each voxel summarises the evidence regarding the existence of an effect at each voxel.

These per-individual images are subsequently combined matching voxel-by-voxel in a second group level linear model, leading to estimates of the overall population effect [Mumford and Nichols, 2006]. Combing the subject data of each voxel independently ignores the between-subject variability and therefore MUA analyses can lead to poor spatial precision or even failure to detect an effect especially when considerable spatial variability is present among the group of participants. To address this limitation, several spatial models have been proposed [Hartvig and

Jensen, 2000; Flandin and Penny, 2007; Bowman *et al.*, 2008; Xu *et al.*, 2009, to name a few] allowing voxels to borrow information between each other and thus are an improvement compared to the standard MUA approach.

However, these methods do not make use of existing information that might be available from previously conducted studies, in order to improve the estimation. This can be of great importance in fMRI studies where sample sizes are typically very small [Carp, 2012] and often therefore suffer from low power [Button *et al.*, 2013]. Exploiting existing knowledge imposed as prior distributions in a Bayesian approach has the potential to improve the power of studies with few participants and detect effects that could otherwise remain undetectable.

In this chapter, we propose a hierarchical Bayesian model for group fMRI studies. Our model is based on the model of Xu *et al.* [2009] but improves upon their approach in the following three aspects. Firstly, our model estimates both activations and deactivations (drops in BOLD response) as opposed to activations only. This is particularly important to facilitate comparison of a task with some condition other than the resting state (baseline). Secondly, we use spatial point process models instead of mixture models at the higher levels of the hierarchy which enables us to improve the computational efficiency of the algorithm used for posterior simulation. Finally, the model uses the output of CBMA meta-analysis as a prior for the effects in the new study in order to gain power.

The chapter is structured as follows. In Section 5.2 we present the details of our Bayesian hierarchical model for group fMRI studies and Section 5.3 discusses posterior inferences for this model. Simulation studies are presented in Section 5.4 and real data analysis in Section 5.5. Section 5.6 contains a summary of our findings and sets the directions of our future work.

## 5.2 The model

The proposed hierarchical model consists of 3 levels. At the bottom level we model the data, a $T$ image for each individual which summarises the evidence for activation/deactivation during the course of the experiment. Voxels that show an effect are assumed to cluster around the individual components that are found in level 2 of the model and are specific to each subject. At the highest level 1 we model the study centers which are the locations in the brain around which effects consistently appear for all participants of a study. We now give the details of the model explaining the 3 different levels of the hierarchy separately.

**Level 3.** Let $y_{iv}$ denote the observed $T$ statistic for subject $i$ at voxel $v$,

$i = 1, \ldots, I$ and $v = 1, \ldots, V$. Following Xu *et al.* [2009], we assume that the data have a mixture distribution with three different normally distributed mixture components. The first component corresponds to the background signal, the second component corresponds to activations and the third to deactivations. The density is:

$$\pi\left(y_{iv} \mid p_{iv0}, m_0, s_0^2, \left\{p_{ijv}, m_{ij}, s_{ij}^2\right\}_{j=1}^{n(\mathbf{x}_i)}\right) = p_{iv0}\phi\left(y_{iv} \mid m_0, s_0^2\right) + \sum_{j=1}^{n\left(\mathbf{x}_i^+\right)} p_{ivj}\phi\left(y_{iv} \mid m_{ij}, s_{ij}^2\right) + \sum_{j=n\left(\mathbf{x}_i^+\right)+1}^{n\left(\mathbf{x}_i^+\right)+n\left(\mathbf{x}_i^-\right)} p_{ivj}\phi\left(y_{iv} \mid m_{ij}, s_{ij}^2\right), \quad (5.1)$$

where $p_{ivj}$ are the mixing weights; $\phi(x \mid \mu, \sigma^2)$ is the probability density function of a $\mathcal{N}(\mu, \sigma^2)$ evaluated at $x$; $m_0, s_0^2$ are the background mean and variance; $\mathbf{x}_i^+$ and $\mathbf{x}_i^-$ are the individual component processes for increases and decreases, respectively, with cardinalities $n(\cdot)$; and $m_{ij}, s_{ij}^2$ are mean and variance for individual component $x_{ij} \in \mathbf{x}_i = \mathbf{x}_i^+ \cup \mathbf{x}_i^-$. Equation (5.1) suggests that the observed $T$ values are either due to random fluctuations in fMRI signal which are represented by the background component or due to some systematic positive or negative effect accounted for by the individual component process.

For the background component we *a priori* assume that for all voxels $p_{iv0} \propto m$, where $m$ is fixed. For the remaining components we assume that the mixing weights are a decreasing function of the distance to a given voxel and in particular that:

$$p_{ivj} \propto \phi_3\left(\boldsymbol{\nu}_v \mid x_{ij}, \mathbf{S}_{ij}\right), \quad j = 1, \ldots, n\left(\mathbf{x}_i^+\right) + n\left(\mathbf{x}_i^-\right),$$

where $\phi_3(x \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of a three-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and $\boldsymbol{\nu}_v$ is the location of the center of voxel $v$. Overall, for all participants $i = 1, \ldots, I$ and voxels $v = 1, \ldots, V$ we have that:

$$p_{ivj} = \begin{cases} \dfrac{m}{m + \sum_{k=1}^{n(\mathbf{x}_i)} \phi_3(\boldsymbol{\nu}_v \mid x_{ik}, \mathbf{S}_{ik})}, & j = 0 \\ \dfrac{\phi_3(\boldsymbol{\nu}_v \mid x_{ij}, \mathbf{S}_{ij})}{m + \sum_{k=1}^{n(\mathbf{x}_i)} \phi_3(\boldsymbol{\nu}_v \mid x_{ik}, \mathbf{S}_{ik})}, & j > 0 \end{cases}. \quad (5.2)$$

The formulation in Equation (5.2) implies that when a voxel is coincident with an individual center $x_{i\ell}$ and no other individual center is near, then the prior probability of the voxel being background is $m/(m + (2\pi)^{-3/2}|\mathbf{S}_{i\ell}|^{-1/2})$. One can thus set $m$ based on prior beliefs. If $\mathbf{S}^*$ is our best prior guess for the spread of effects $\mathbf{S}_{ij}$ and $q$ is the expected proportion of voxels that are not activated by the task under

investigation, then:

$$m = \frac{q}{1-q}(2\pi)^{-3/2}|\mathbf{S}^*|^{-1/2}. \tag{5.3}$$

In our applications we set $m$ using Equation (5.3). For $\mathbf{S}^*$ we use the mode of the prior on $\mathbf{S}_{ij}$ (defined in the following section) and we set $q$ to some high value, either 0.95 or 0.99 since fMRI experiments are typically designed to activate only a small proportion of the voxels.

When no effect is present, the $t$ statistics should follow the nominal null distribution. To reflect this belief we assume that $m_0 \sim \mathcal{N}(0,1)$ and $s_0^2 \sim \text{IG}(10^{-3}, 10^{-3})$. Prior specifications for $x_{ij}, \mathbf{S}_{ij}, m_{ij}$ and $s_{ij}^2$ are given at level 2.

**Level 2.** At level 2 we model the individual components for increases $\mathbf{x}_i^+ = \left\{x_{ij}\right\}_{j=1}^{n(\mathbf{x}_i^+)}$ and decreases $\mathbf{x}_i^- = \left\{x_{ij}\right\}_{j=n(\mathbf{x}_i^+)+1}^{n(\mathbf{x}_i^+)+n(\mathbf{x}_i^-)}$. These are the locations in each individual's brain around which task effects are present. Individual components cluster around the study centers. We assume that individual components are realisations from point processes $\mathbf{X}_i^+$ and $\mathbf{X}_i^-$ which are obtained as following.

For increases, we associate with each study center $z_k \in \mathbf{z}^+$ a Poisson process $\mathbf{X}_{iz_k}^+$ of offspring normally distributed around $z_k$ with covariance $\mathbf{\Sigma}_{z_k}$, $k = 1, \ldots, n(\mathbf{z}^+)$. Then, $\mathbf{X}_i^+$ is obtained as the union of the offspring:

$$\mathbf{X}_i^+ = \bigcup_{z_k \in \mathbf{z}^+} \mathbf{X}_{iz_k}^+. \tag{5.4}$$

This analytically tractable model is also known as the Cox cluster process. The intensity function of the cluster process $\mathbf{X}_i^+$, $\rho_i^+$ is given as:

$$\rho_i^+\left(\xi \mid \epsilon, \eta, \{z_k, \mathbf{\Sigma}_{z_k}\}_{k=1}^{n(\mathbf{z}^+)}\right) = \epsilon + \eta \sum_{z_k \in \mathbf{z}^+} \phi_3\left(\xi \mid z_k, \mathbf{\Sigma}_{z_k}\right), \tag{5.5}$$

where $\frac{\eta}{\int_{\mathcal{B}} \phi_3(\xi | z_k, \mathbf{\Sigma}_{z_k}) d\xi}$ is the expected number of offspring of $z_k \in \mathbf{z}^+$ and $\epsilon$ is introduced for idiosyncratic components that do not cluster around any study centers. For the purposes of this work we keep $\epsilon$ and $\eta$ fixed and set their values based on application. The reason is that the spatial Poisson model offers considerable flexibility even for fixed $\eta$ and $\epsilon$; it is however possible to perform inference on these parameters as well.

Additionally, we associate with each $x_{ij} \in \mathbf{x}_i^+, j = 1, \ldots, n\left(\mathbf{x}_i^+\right)$, a set of marks $\left\{\mathbf{S}_{ij}, m_{ij}, s_{ij}^2\right\}$, that reflect the spread, the mean and the variability of the effect. These marks are independent of the locations $x_{ij}$ and have the following prior distributions. For the spreads, we assume that $\mathbf{S}_{ij} \sim \text{IW}\left(\mathbf{T}_{\mathbf{x}}, d_{\mathbf{x}}\right)$. The degrees of freedom are set to $d_{\mathbf{x}} = 10$. As discussed before, we choose $\mathbf{T}_{\mathbf{x}}$ to reflect our

prior beliefs regarding the spread of the effects around individual components. For increases, the mean of the effect should have a positive expectation and hence $m_{ij} \sim \mathcal{N}_{(0,+\infty)}\left(\mu^+, \sigma^2_{m^+}\right)$. Hyperpriors reflect our prior beliefs regarding the effect size and variance: $\mu^+ \sim \mathcal{N}(3, 10)$ and $\sigma^2_{m^+} \sim \text{IG}\left(10^{-3}, 10^{-3}\right)$. Finally, we assign $s^2_{ij}$ a $\text{IG}\left(2, \beta_{s^2_+}\right)$ prior. This prior is chosen because it has an infinite second moment and therefore allows for the presence of outlier values. To take advantage of the conjugacy, we assume that $\beta_{s^2_+} \sim \text{Ga}\left(10^{-2}, 10^{-2}\right)$. Overall, the joint distribution of $\mathbf{x}^+_i$ and the marks is:

$$\pi\Big(\left\{x_{ij}, \mathbf{S}_{ij}, m_{ij}, s_{ij}\right\}_{j=1}^{n\left(\mathbf{x}^+_i\right)} \mid \epsilon, \eta, \left\{z_k, \mathbf{\Sigma}_{z_k}\right\}_{k=1}^{n\left(\mathbf{z}^+\right)}, \mathbf{T}_{\mathbf{x}}, d_{\mathbf{x}}, \mu^+, \sigma^2_{m^+}, \beta_{s^2_+}\Big) =$$

$$\exp\left(|\mathcal{B}| - \int_{\mathcal{B}} \rho^+_i\left(\xi \mid \epsilon, \eta, \left\{z_k, \mathbf{\Sigma}_{z_k}\right\}_{k=1}^{n\left(\mathbf{z}^+\right)}\right) d\xi\right) \prod_{j=1}^{n\left(\mathbf{x}^+_i\right)} \left[\rho^+_i\left(x_{ij} \mid \epsilon, \eta, \left\{z_k, \mathbf{\Sigma}_{z_k}\right\}_{k=1}^{n\left(\mathbf{z}^+\right)}\right)\right.$$

$$\left. \pi\left(\mathbf{S}_{ij} \mid \mathbf{T}_{\mathbf{x}}, d_{\mathbf{x}}\right) \pi\left(m_{ij} \mid \mu^+, \sigma^2_{m^+}\right) \pi\left(s^2_{ij} \mid \beta_{s^2_+}\right)\right]. \quad (5.6)$$

The individual component process for decreases is defined in similar fashion. Each $z_k \in \mathbf{z}^-$ is associated with a set of points $\mathbf{X}^-_{iz_k}$ which are normally distributed around $z_k$ with covariance $\mathbf{\Sigma}_{z_k}$. The independent cluster process $\mathbf{X}^-_i$ is obtained as:

$$\mathbf{X}^-_i = \bigcup_{z_k \in \mathbf{z}^-} \mathbf{X}^-_{iz_k}, \quad (5.7)$$

and has intensity:

$$\rho^-_i\left(\xi \mid \epsilon, \eta, \left\{z_k, \mathbf{\Sigma}_{z_k}\right\}_{k=n\left(\mathbf{z}^+\right)+1}^{n\left(\mathbf{z}^+\right)+n\left(\mathbf{z}^-\right)}\right) = \epsilon + \eta \sum_{z_k \in \mathbf{z}^-} \phi_3\left(\xi \mid z_k, \mathbf{\Sigma}_{z_k}\right). \quad (5.8)$$

For the marks $\left\{\mathbf{S}_{ij}, m_{ij}, s^2_{ij}\right\}_{j=n\left(\mathbf{x}^+_i\right)+1}^{n\left(\mathbf{x}^+_i\right)+n\left(\mathbf{x}^-_i\right)}$ we use similar prior specifications as for increases. In particular, $\mathbf{S}_{ij} \sim \text{IW}\left(\mathbf{T}_{\mathbf{x}}, d_{\mathbf{x}}\right)$, $m_{ij} \sim \mathcal{N}_{(-\infty,0)}\left(\mu^-, \sigma^2_{m^-}\right)$ and $s_{ij} \sim \text{IG}\left(2, \beta_{s^2_-}\right)$. Hyperpriors are defined as following: $\mu^- \sim \mathcal{N}(-3, 10)$, $\sigma^2_{m^-} \sim \text{IG}$

$(0.001, 0.001)$ and $\beta_{s^2_-} \sim \text{Ga}\left(10^{-2}, 10^{-2}\right)$. The joint density of points and marks is:

$$\pi\Big(\left\{x_{ij}, \mathbf{S}_{ij}, m_{ij}, s_{ij}\right\}_{j=n\left(\mathbf{x}_i^+\right)+1}^{n\left(\mathbf{x}_i^+\right)+n\left(\mathbf{x}_i^-\right)} \mid \epsilon, \eta, \left\{z_k, \mathbf{\Sigma}_{z_k}\right\}_{k=n\left(\mathbf{z}^+\right)+1}^{n\left(\mathbf{z}^+\right)+n\left(\mathbf{z}^-\right)}, \mathbf{T}_{\mathbf{x}}, d_{\mathbf{x}}, \mu^-, \sigma^2_{m^-}, \beta_{s^2_-}\Big)$$
$$= \exp\left(|\mathcal{B}| - \int_{\mathcal{B}} \rho_i^-\left(\xi \mid \epsilon, \eta, \left\{z_k, \mathbf{\Sigma}_{z_k}\right\}_{k=n\left(\mathbf{z}^+\right)+1}^{n\left(\mathbf{z}^+\right)+n\left(\mathbf{z}^-\right)}\right) d\xi\right)$$
$$\prod_{j=n\left(\mathbf{x}_i^+\right)+1}^{n\left(\mathbf{x}_i^+\right)+n\left(\mathbf{x}_i^-\right)} \Bigg[\rho_i^-\left(x_{ij} \mid \epsilon, \eta, \left\{z_k, \mathbf{\Sigma}_{z_k}\right\}_{k=n\left(\mathbf{x}_i^+\right)+1}^{n\left(\mathbf{z}^+\right)+n\left(\mathbf{x}_i^-\right)}\right)$$
$$\pi\left(\mathbf{S}_{ij} \mid \mathbf{T}_{\mathbf{x}}, d_{\mathbf{x}}\right) \pi\left(m_{ij} \mid \mu^-, \sigma^2_{m^-}\right) \pi\left(s^2_{ij} \mid \beta_{s^2_-}\right)\Bigg]. \quad (5.9)$$

**Level 1.** At the highest level, we model the study centers $\mathbf{z} = \mathbf{z}^+ \cup \mathbf{z}^-$. These are the centers of brain regions that are consistently associated with increases or decreases in BOLD signal among participants. We assume that $\mathbf{z}^+$ and $\mathbf{z}^-$ are realisations of Cox processes $\mathbf{Z}^+$ and $\mathbf{Z}^-$ driven by random intensities $\boldsymbol{\lambda}^+$ and $\boldsymbol{\lambda}^-$ respectively.

At the absence of prior information we can assume homogenous $\lambda^+$ and $\lambda^-$ and that $\lambda^+ \mid \alpha_{\lambda^+}, \beta_{\lambda^+} \sim \text{Ga}\left(\alpha_{\lambda^+}, \beta_{\lambda^+}\right)$ and $\lambda^- \mid \alpha_{\lambda^-}, \beta_{\lambda^-} \sim \text{Ga}\left(\alpha_{\lambda^-}, \beta_{\lambda^-}\right)$. However, when CBMA data are available we can use it as prior information in the model. In such cases we have $\left\{\mathbf{w}_k^+\right\}_{k=1}^K$ and $\left\{\mathbf{w}_k^-\right\}_{k=1}^K$, the foci associated with increases and decreases in signal respectively, as obtained from $K$ previously conducted studies. There are several ways in which the information that $\left\{\mathbf{w}_k^+\right\}_{k=1}^K$ and $\left\{\mathbf{w}_k^-\right\}_{k=1}^K$ provide can be incorporated into the model, some of which are discussed in the following section.

Attached to each $z \in \mathbf{z} = \mathbf{z}^+ \cup \mathbf{z}^-$ is a mark $\mathbf{\Sigma}_z$ which characterises the spread of individual centers around $z$. *A priori* $\mathbf{\Sigma}_z \sim \text{IW}\left(\mathbf{T}_{\mathbf{z}}, d_{\mathbf{z}}\right)$, where $d_{\mathbf{z}} = 10$ and $\mathbf{T}_{\mathbf{z}}^{-1} \sim \text{IW}\left(\mathbf{T}_0, d_0\right)$ with $\mathbf{T}_0$ being fixed based on prior beliefs regarding the spread of individual components around the study centers and $d_0 = 10$. Overall, the joint density of $\mathbf{z}^+$ and marks is:

$$\pi\Big(\left\{z_k, \mathbf{\Sigma}_{z_k}\right\}_{k=1}^{n\left(\mathbf{z}^+\right)} \mid \boldsymbol{\lambda}^+, \mathbf{T}_{\mathbf{z}}, d_{\mathbf{z}}\Big) = \exp\left(|\mathcal{B}| - \int_{\mathcal{B}} \boldsymbol{\lambda}^+\left(\xi \mid \left\{\mathbf{w}_k^+\right\}_{k=1}^K\right) d\xi\right)$$
$$\prod_{k=1}^{n\left(\mathbf{z}^+\right)} \Big[\boldsymbol{\lambda}^+\left(z_k \mid \left\{\mathbf{w}_k^+\right\}_{k=1}^K\right) \pi\left(\mathbf{\Sigma}_{z_k} \mid \mathbf{T}_{\mathbf{z}}, d_{\mathbf{z}}\right)\Big], \quad (5.10)$$

whereas for $\mathbf{z}^-$ it is:

$$\pi\Big(\{z_k, \boldsymbol{\Sigma}_{z_k}\}_{k=n(\mathbf{z}^+)+1}^{n(\mathbf{z}^+)+n(\mathbf{z}^-)} \mid \boldsymbol{\lambda}^-, \mathbf{T}_{\mathbf{z}}, d_{\mathbf{z}}\Big) = \exp\Big(|\mathcal{B}| - \int_{\mathcal{B}} \boldsymbol{\lambda}^-\Big(\xi \mid \{\mathbf{w}_k^-\}_{k=1}^K\Big)\, d\xi\Big) \prod_{k=n(\mathbf{z}^+)+1}^{n(\mathbf{z}^+)+n(\mathbf{z}^-)} \Big[\boldsymbol{\lambda}^-\Big(z_k \mid \{\mathbf{w}_k^-\}_{k=1}^K\Big)\, \pi\left(\boldsymbol{\Sigma}_{z_k} \mid \mathbf{T}_{\mathbf{z}}, d_{\mathbf{z}}\right)\Big]. \quad (5.11)$$

Figure 5.1 is a graphical representation of the model for a study with $I = 3$ participants.

## 5.3 Posterior inferences

The posterior distribution of the hierarchical model described in Section 5.2 is shown in Appendix B.1. This posterior cannot be computed analytically and therefore we need to resort to Markov chain Monte Carlo to obtain samples from it. We propose a hybrid Gibbs sampler where parameters are drawn from their full conditional distributions, either one at a time or jointly in blocks of several parameters. In this section we summarise the techniques used for posterior simulations. For the full details of the algorithm, see Section B.2 of the Appendix.

The main challenge for posterior sampling in our model is that the total number of individual components as well as the total number of study centers is unknown and hence needs to be estimated. In their model, Xu *et al.* [2009] use the *reversible jump* MCMC algorithm [Green, 1995] to estimate the total number of mixing components in their model. We instead adopt the *spatial birth-and-death* [Preston, 1977; Møller and Waagepetersen, 2004; Huber, 2011] algorithm because it is easier to design efficient trans-dimensional proposals for spatial point processes. The spatial birth-and-death process is continuous time Markov chain whose moves are either births or deaths and can be used to simulate spatial point processes. We now explain the general procedure.

Assume that we wish to simulate from a point process $\mathbf{A}$ on $\mathcal{B}$ with density $\pi\big(\mathbf{a} \mid \mathbf{u}\big)$, where $\mathbf{u}$ are the data. Let $b\left(\xi\right)$ be the birth rate for a new point $\xi$ and $d\left(\xi\right)$ be the death rate for removing a point $\xi \in \mathbf{a}$; it was shown by Preston [1977] that if these rates satisfy the detailed balance equation

$$\pi\left(\mathbf{a} \mid \mathbf{u}\right) b\left(\xi\right) = \pi\left(\mathbf{a} \cup \{\xi\} \mid \mathbf{u}\right) d\left(\xi\right), \quad (5.12)$$

then the chain has unique equilibrium distribution $\pi\big(\mathbf{a} \mid \mathbf{u}\big)$ to which it will converge regardless of the initial state. Given the current state $\mathbf{a}$, the waiting times until a

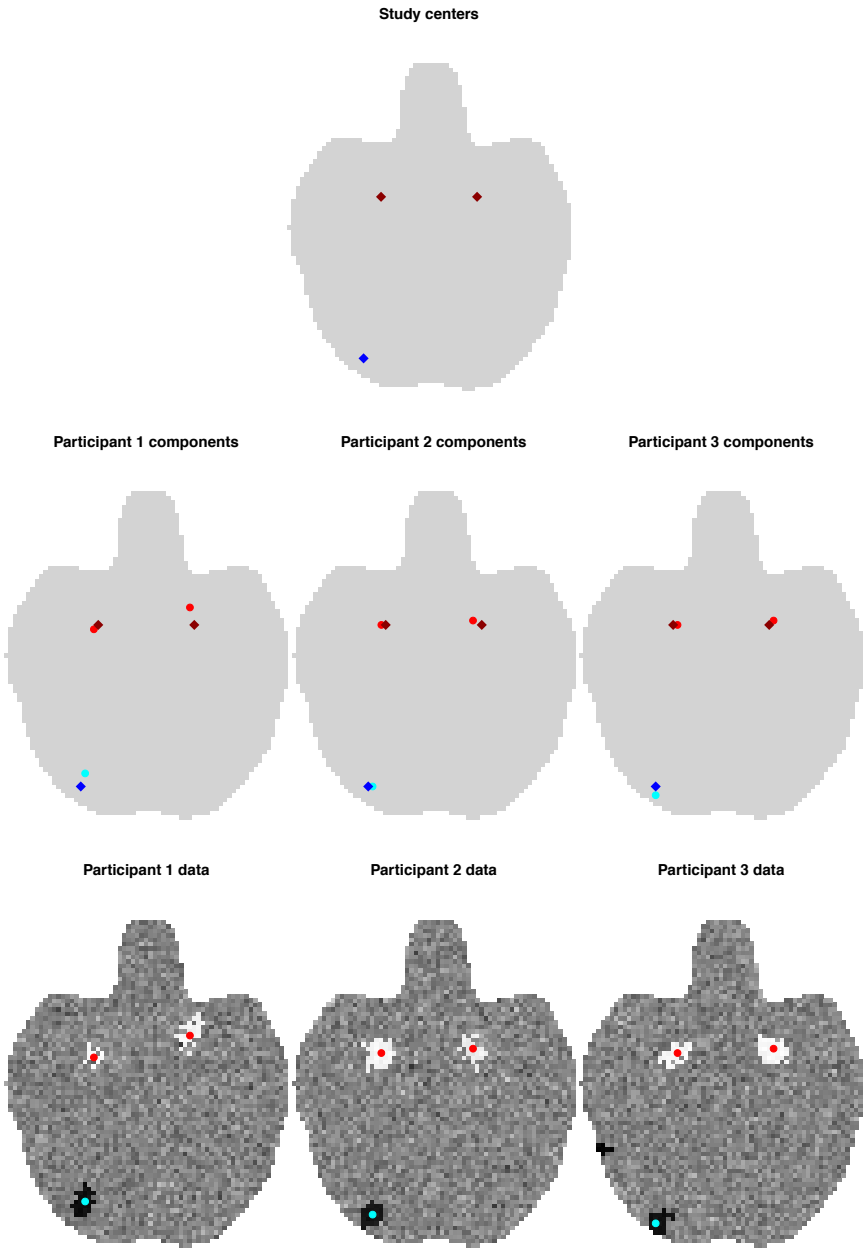



Figure 5.1: Realisation of the group fMRI model for 3 participants. At level 1 (top) we have the study centers for increases ($\mathbf{z}^+$, red rhombuses) and decreases ($\mathbf{z}^-$, blue rhombus). At level 2 we have individual components for increases ($\mathbf{x}_i^+$, light red circles) and decreases ($\mathbf{x}_i^-$, light blue circles). Finally, at the bottom level 3 we have the $T$ images for each participant. Bright colours indicate voxels with high positive values whereas dark colours indicate voxels with low negative values.

new event are exponentially distributed with rate $B + D$, where $B = \int_{\mathcal{B}} b(\xi)\, d\xi$ and $D = \sum_{\xi \in \mathbf{a}} d(\xi)$ are the total rates; conditional on an event a birth occurs with probability $B/(B + D)$ and a death with probability $D/(B + D)$. For a birth, we draw a new point from the density $b(\xi)/B$ whereas for a death a point $\xi \in \mathbf{a}$ is chosen with probability $d(\xi)/D$.

The spatial birth-and-death algorithm can be particularly useful in the context of hierarchical models where $\pi(\mathbf{a} \mid \mathbf{u})$ is the full conditional of the process $\mathbf{a}$ that we wish to simulate from and $\mathbf{u}$ can be viewed as the observed data or parameters at lower levels of the hierarchical model. As a result, one can use $\mathbf{u}$ to design efficient birth rates so that new points are drawn in the vicinity of the data [van Lieshout and Baddeley, 2002]. Then, the death rates can be found based on the detailed balance Equation (5.12). The total running time $T$ of the algorithm is a tuning parameter and needs to be specified so that there is a good balance between the mixing properties of the chain and the computational effort required; a common choice is to set $T = 1/B$ [Kang *et al.*, 2011]. Pseudocode for the spatial birth-and-death algorithm can be found in Algorithm 1.

**input** : $\mathbf{a}$, $T$, $b(\xi)$, $d(\xi)$
**output**: $\mathbf{a}^*$

$B \leftarrow \int_{\mathcal{B}} b(\xi)\, d\xi$;
$D \leftarrow \sum_{\xi \in \mathbf{a}} d(\xi)$;
Draw $t \sim \exp(B + D)$;
**while** $t \leq T$ **do**
  Draw $u \sim \text{Uni}[0, 1]$;
  **if** $u \leq \frac{B}{B+D}$ **then**
    Draw $\xi \sim \frac{b}{B}$;
    $\mathbf{a} \leftarrow \mathbf{a} \cup \xi$;
  **end**
  **else**
    Draw $\xi \in \mathbf{a}$ w.p. $\frac{d(\xi)}{D}$;
    $\mathbf{a} \leftarrow \mathbf{a} \setminus \xi$;
  **end**
  $B \leftarrow \int_{\mathcal{B}} b(\xi)\, d\xi$;
  $D \leftarrow \sum_{\xi \in \mathbf{a}} d(\xi)$;
  $t \leftarrow t + \exp(B + D)$
**end**
$\mathbf{a}^* \leftarrow \mathbf{a}$

**Algorithm 1:** Spatial birth-and-death algorithm.

We use the spatial birth-and-death algorithm to update the total number of

increase and decrease components per individual, $n\left(\mathbf{x}_i^+\right)$ and $n\left(\mathbf{x}_i^-\right)$, as well as the study centers for increases and decreases, $n\left(\mathbf{z}^+\right)$ and $n\left(\mathbf{z}^-\right)$. Once these are updated, we can use standard techniques to update the remaining model parameters. Full conditional distributions can be found in closed form for the following: $m_0$, $s_0^2$, $m_{ij}$, $s_{ij}^2$, $\mu^+$, $\mu^-$, $\sigma_{m^+}^2$, $\sigma_{m^-}^2$, $\beta_{s_+^2}$, $\beta_{s_-^2}$, $\mathbf{T}_{\mathbf{z}}$, for all $i = 1, \ldots, I$ and $j = 1, \ldots, n\left(\mathbf{x}_i\right)$. Thus, these parameters are simulated with Gibbs steps directly from their full conditional distributions. For the remaining parameters, $\mathbf{x}_{ij}$, $\mathbf{S}_{ij}$, $\mathbf{z}_k$ and $\mathbf{\Sigma}_{z_k}$, where $i = 1, \ldots, I$, $j = 1, \ldots, n\left(\mathbf{x}_i\right)$ and $k = 1, \ldots, n\left(\mathbf{z}\right)$, we use simple Metropolis-Hastings updates as demonstrated in Appendix B.2.

There are several ways in which we can handle the update of study center intensities $\boldsymbol{\lambda}^+$ and $\boldsymbol{\lambda}^-$. With no CBMA data, one may assume constant rates $\lambda^+$ and $\lambda^-$, and then either keep these parameters fixed throughout posterior simulation or update at each iteration of the MCMC with a Metropolis-Hastings step. With CBMA data, one option is to run a point process CBMA model on the foci $\left\{\mathbf{w}_k^+\right\}_{k=1}^K$ and $\left\{\mathbf{w}_k^-\right\}_{k=1}^K$, using subsequently the posterior distributions $\pi\Big(\boldsymbol{\lambda}^+ \mid \left\{\mathbf{w}_k^+\right\}_{k=1}^K\Big)$ and $\pi\Big(\boldsymbol{\lambda}^- \mid \left\{\mathbf{w}_k^-\right\}_{k=1}^K\Big)$ to obtain point estimates $\hat{\boldsymbol{\lambda}}^+$ and $\hat{\boldsymbol{\lambda}}^-$ and ultimately set $\boldsymbol{\lambda}^+ = \hat{\boldsymbol{\lambda}}^+$ and $\boldsymbol{\lambda}^- = \hat{\boldsymbol{\lambda}}^-$. We employ this strategy for our simulation studies. Another option, the one that we use for our real data applications, would be to marginalise over the CBMA posteriors that is: save $L$ meta-analysis draws $\left\{\boldsymbol{\lambda}^{+(\ell)}, \boldsymbol{\lambda}^{-(\ell)}\right\}_{\ell=1}^L$ and then at each iteration of the MCMC for the group fMRI model uniformly draw the study center intensities from these $L$ candidate values. Finally, it is also possible to set $\mathbf{w}_{K+1}^+ = \mathbf{z}^+$ and $\mathbf{w}_{K+1}^- = \mathbf{z}^-$ and update CBMA intensities within the group fMRI sampler, with an additional step using for example the LGCP model presented in Chapter 4.

Several quantities of interest can be obtained through the posterior distribution of the model parameters. For participant $i$ and voxel $v$ we introduce the allocation parameter $\delta_{iv}$ indicating which component the voxel belongs to. Using $\delta_{iv}$ we can estimate the marginal posterior probability that a voxel is activated as the proportion of posterior samples for which $\delta_{iv}$ points to an increase component $x_{ij} \in \mathbf{x}_i^+$; the probability that a voxel is associated with decreases in BOLD signal can be estimated in similar fashion. For each individual, we approximate the voxel-wise probability that an individual component is located within the voxel $v$ as $\frac{1}{L}\sum_{\ell=1}^{L}\sum_{j=1}^{n(\mathbf{x}_i^{(\ell)})}\mathbf{1}_{\left\{x_{ij}^{(\ell)} \in v\right\}}$, where $\mathbf{x}_i^{(\ell)}$ is the $\ell$-th posterior draw from the individual component process of participant $i$ and $L$ is the total number of MCMC iterations. For study centers, we can similarly estimate the probability that there is a focus in any voxel $v$ as $\frac{1}{L}\sum_{\ell=1}^{L}\sum_{k=1}^{n(\mathbf{z}^{(\ell)})}\mathbf{1}_{\left\{z_k^{(\ell)} \in v\right\}}$, where $\mathbf{z}_k^{(\ell)}$ is the $\ell$-th posterior draw

from the study center process. Note that for both $\mathbf{x}_i$ and $\mathbf{z}$ it is also possible to construct posterior intensity functions using the generated points at each iteration. This facilitates comparison between individuals and comparison of study centers with previous CBMA findings. All these inferences are very hard to obtain with massively univariate approaches and this is therefore an advantage of the model compared to these methods. Examples are given in the following sections.

## 5.4 Simulation studies

In this section we perform a series of simulation studies to assess the performance of our model with synthetic data. We use the same brain atlas as in Chapter 4 which consists of 216,040 $2\text{mm}^3$ voxels. In Section 5.4.1 we describe the mechanism under which data are generated, and in Section 5.4.2 we validate our model for one of the generated datasets and demonstrate the inferences that can be obtained with our approach. In Section 5.4.3 we study the sensitivity of the method to the specification of the background probability parameter $m$.

### 5.4.1 Simulation setup

Data are generated as follows. We assume that there are a total of 8 study centers: 4 of them are associated with increases in BOLD signal and the remaining 4 are associated with decreases. The locations of the centers are chosen based on the meta-analysis of emotion and executive control studies that was presented in Section 4.5. In particular, we choose $\mathbf{z}^+ = \{z_k\}_{k=1}^4$ to be the centers of 4 regions around which voxels were found to have high posterior intensities for emotions; for $\mathbf{z}^- = \{z_k\}_{k=5}^8$ we use executive control posterior intensities. The locations are shown in Table 5.1. Each one of the study centers is associated with a different mean effect magnitude $\mu_k$, $k = 1, \ldots, 8$, in order to cover scenarios where the observed $T$ statistics are low, moderate or high in absolute value. Each center is further assigned a different $\sigma_k^2$ which corresponds to the variability of the data around $\mu_k$. The spatial spread $\mathbf{\Sigma}_k$ of individual centers around the study centers also depends on $k$ with various extents considered. The exact specifications of $\mu_k$, $\sigma_k^2$ and $\mathbf{\Sigma}_k$ can be found in Table 5.1.

The total number number of participants is set to $I = 8$, which is small in order to mimic studies with low power. For each individual $i$, every center $z_k$ produces a single individual component and so all participants have exactly 8 components, 4 for increases and 4 for decreases. The locations of the components, $x_{ik}$, are drawn from a $\mathcal{N}(z_k, \mathbf{\Sigma}_k)$ distribution. Eventually, we simulate the $T$ images conditional on the individual components. For each component $x_{ik}$ we find the voxel $v_{x_{ik}}$ that

| Center | Location | | | $\mu$ | $\sigma^2$ | $\Sigma(\times \mathbf{I_3})$ |
|---|---|---|---|---|---|---|
| $z_1$ | -44 | -80 | -6 | 3.75 | 0.15 | 8 |
| $z_2$ | -20 | -6 | -20 | 2.50 | 0.10 | 2 |
| $z_3$ | 24 | -6 | 20 | 3.00 | 0.20 | 4 |
| $z_4$ | 50 | -74 | -6 | 3.25 | 0.10 | 6 |
| $z_5$ | 0 | 6 | 52 | -2.75 | 0.08 | 3 |
| $z_6$ | -40 | 0 | 36 | -4.00 | 0.10 | 10 |
| $z_7$ | -26 | -70 | 36 | -3.25 | 0.15 | 8 |
| $z_8$ | 38 | -58 | 44 | -3.00 | 0.10 | 5 |

Table 5.1: Summary of our simulation setup for study centers. The locations are determined based on the meta-analysis of emotion and executive control studies that was presented in Section 4.5. Each center is assigned a mean effect size ($\mu$), an effect variability ($\sigma^2$) and a variance covariance mark ($\mathbf{\Sigma}$) that represents the spatial spread of individual centers around it.

contains $x_{ik}$ and then activate B voxels that are close to $v_{x_{ik}}$. B is any integer from 55 to 75 with equal probability and so we select 65 voxels on average. The probability that a voxel $v$ is activated is proportional to $\phi_3\left(\boldsymbol{\nu}_v \mid \boldsymbol{\nu}_{v_{x_{ik}}}, \mathbf{R}_{ik}\right)$, where $\boldsymbol{\nu}_v$ is the location of the center of the voxel $v$ and $\mathbf{R}_{ik} \sim \text{IW}\left(125 \times \mathbf{I}_3, 10\right)$. The inverse Wishart is set such that its mode is near $9 \times \mathbf{I}_3$ and so we expect to observe normal isotropic effects with a standard deviation of 3mm, or 7mm FWHM. Selected voxels around $x_{ik}$ are assigned values from $\mathcal{N}\left(\mu_k, \sigma_k^2\right)$ independently one of another. Voxels that are not associated with any individual component are given values from the standard normal distribution, again independently one of another. An example of a simulated $T$ image for some axial slices can be found in Figure 5.2.

### 5.4.2 Analysis of a single dataset

We generate one dataset using the setup described in Section 5.4.1 and fit our hierarchical model with the following specifications. We set $\epsilon = 2/|\mathcal{B}|$ to allow for an average 2 components that do not cluster around any focus. The expected number of components per study center is set to $\eta = 2$. We *a priori* assume that the effects are isotropic and have 6mm FWHM for each dimension; so we set $\mathbf{T_x} = \frac{14 \times 36}{8 \log 2}\mathbf{I}_3$ which leads $\mathbf{S}_{ij}$ to have the desired mode [1]. To reflect our belief that the spatial variability of components around the foci is higher than the variability of effects around components, we let $\mathbf{T}_0 = \frac{5}{30}$ which a leads to $\mathbb{E}\left[\mathbf{\Sigma}_k\right] = 10$. We fix $\boldsymbol{\lambda}^+$ and $\boldsymbol{\lambda}^-$ to some draw from the posterior distribution of emotion and executive control

[1] The mode of a 3-dimensional IW $(\mathbf{\Sigma}, d)$ distribution is $\mathbf{\Sigma}/(d+4)$.

studies, respectively, as obtained from the application of Section 4.5. Both values are chosen uniformly at random and are re-normalised so that the expected number of points is 4. The algorithm runs for 65,000 iterations, of which we discard the first 15,000 as a burn-in. The remaining 50,000 are used to obtain point estimates, after we apply a thinning factor of 10 for a total of 5,000 posterior draws. We now summarise the results.

Figures 5.2 and 5.3 shows results for participants 1 and 8 of the simulated dataset, for two axial slices. In the first column we show the individual component process on top of the data, as it was at iteration 65,000 of the simulation. We see that the algorithm places components in regions of high signal intensity, as we expected. Note that some activation regions require more than one component to be modelled due to non-Gaussian shape, e.g. activation near the right amygdala on the top row for participant 1. The posterior probabilities of activation and deactivation seem to match the observed signal patterns and are close to 1 for voxels that show an effect. Note that when the signal is weak, for example in the left amygdala activation of participant 8, the posterior probabilities of activation that are assigned are lower.

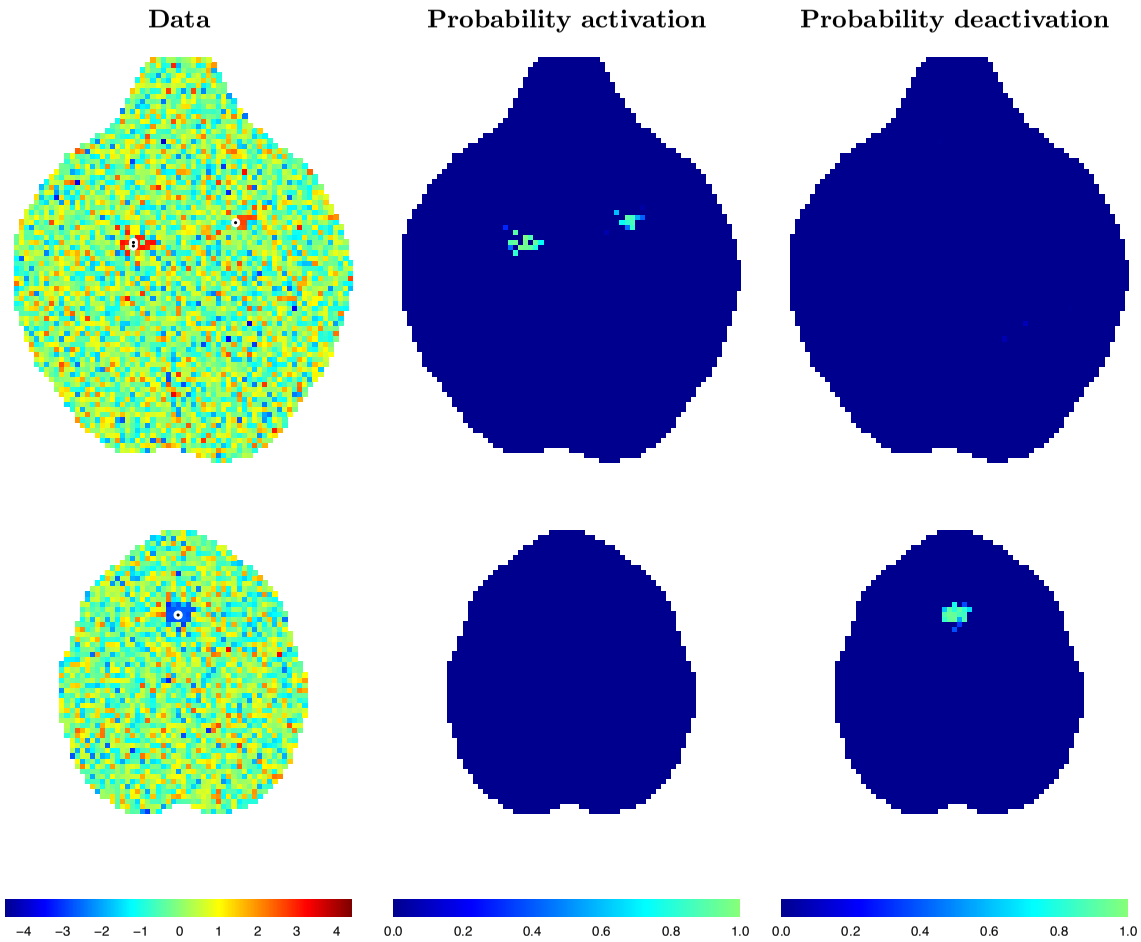



Figure 5.2: Results for participant 1 in the simulation of Section 5.4.2. The first column shows the data and the individual component process (black dots) at iteration 65,000 of the algorithm. The second and third column shows the marginal posterior probabilities of activation and deactivation respectively, as obtained from 5,000 posterior samples. The top row corresponds to axial slice $z = -20$ whereas the bottom row corresponds to axial slice $z = 60$.

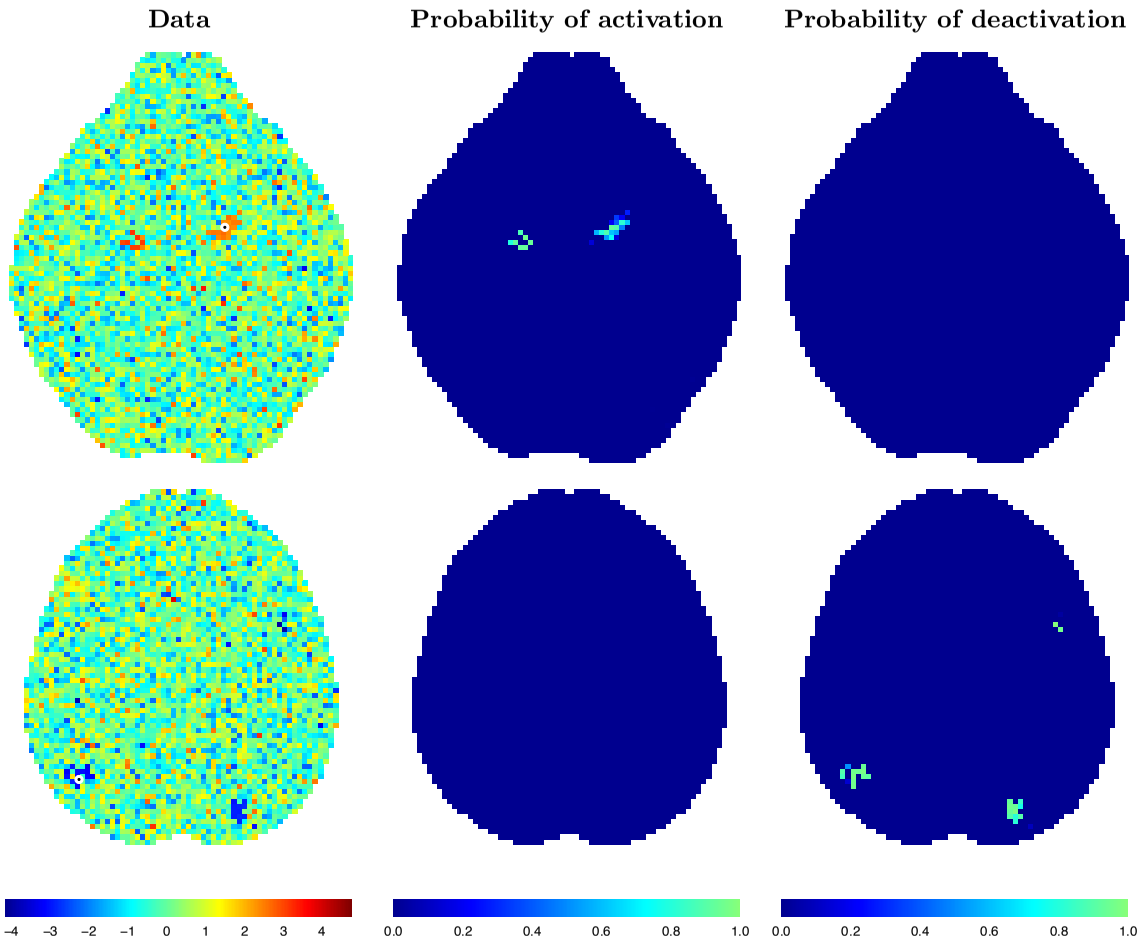



Figure 5.3: Results for participant 8 in the simulation of Section 5.4.2. The first column shows the data and the individual component process (black dots) at iteration 65,000 of the algorithm. The second and third row shows the marginal posterior probabilities of activation and deactivation respectively, as obtained from 5,000 posterior samples. The top row corresponds to axial slice $z = -18$ whereas the bottom row corresponds to axial slice $z = 44$.

Figure 5.4 shows the total number of individual increase and decrease components for participant 1 (top left and right panels respectively) as well as participant 8 (middle left and right panels respectively). We see that the count processes reach their stationary distributions rapidly and exhibit fairly low autocorrelation after that. Another interesting observation is that the model assigns roughly 8 increase and decrease components per individual, double the number we used to simulate the data. Again, we can assume that this stems from some non-Gaussian shapes in the simulated activations but also the strict prior on the background probability parameter $m$, as we discuss in the following section. The total number of study centers (Figure 5.4, bottom panel) also reaches the posterior mode quickly and demonstrates good mixing properties. The posterior medians are 4 and 5 respectively, very close to the true values.

We also calculate the marginal posterior probabilities that there is a study center at each voxel and plot results for some of the axial slices where we have placed the effects. These can be found in Figure 5.6. We generally see that the model captures the presence of an effect near almost all study centers. The locations of the study centers are given in the bottom panel of the Figure. Note that the effects for which we assigned the least spatial variability of individual components around them also have the best accuracy. An example is the effects near the right and left amygdala, the bilateral blobs in the first column of Figure 5.6. On the contrary, the effects at slice $z = +36$ for decreases which had the bigger component variability are the ones for which the posterior is less concentrated.

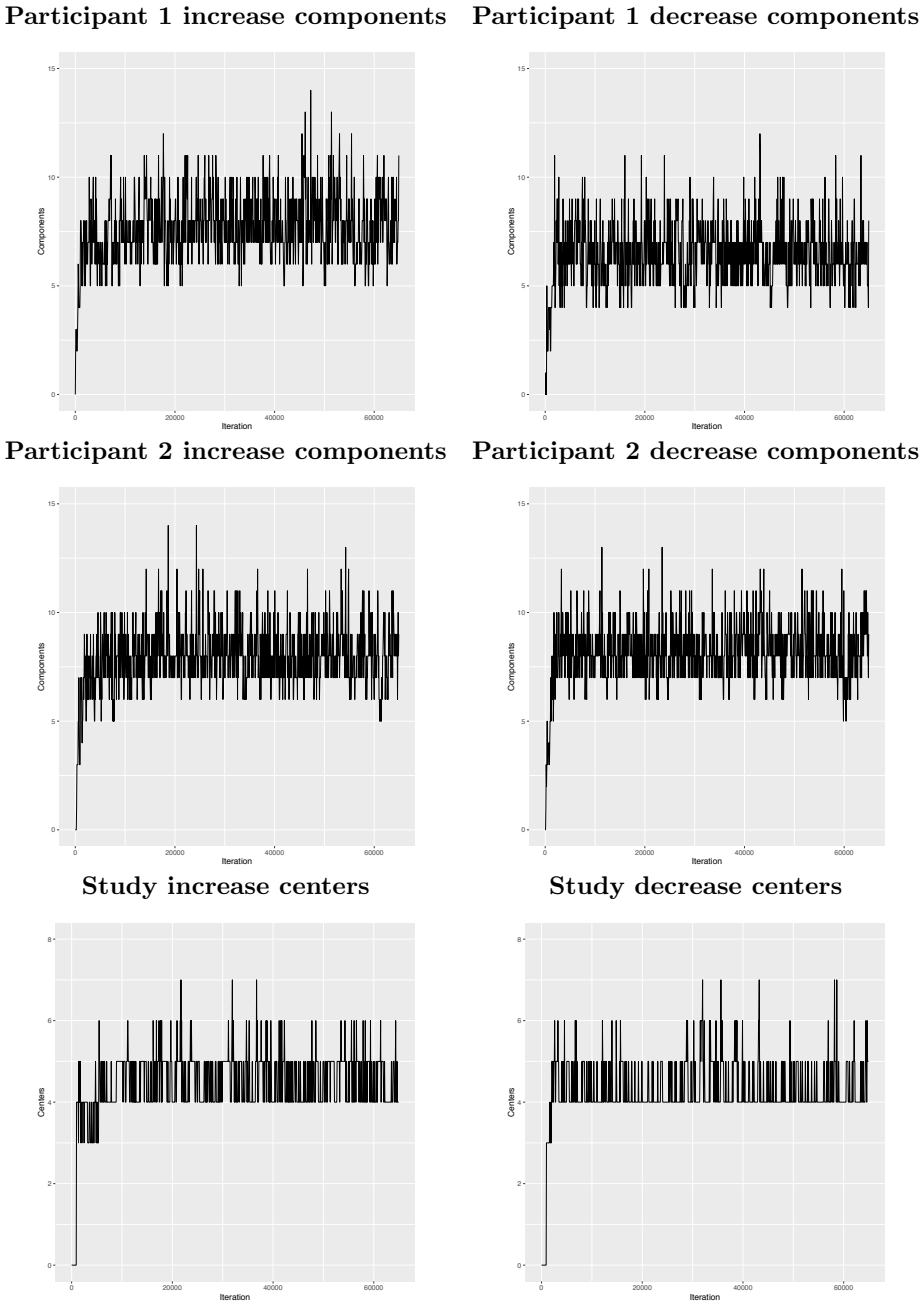



Figure 5.4: Results for the simulated dataset of Section 5.4.2. The figure shows posterior traceplots for the total number of individual components of participant 1 (top row), the total number of individual components of participant 8 (middle row) and the total number of study centers (bottom row). The left column corresponds to increases whereas the right column corresponds to decreases.

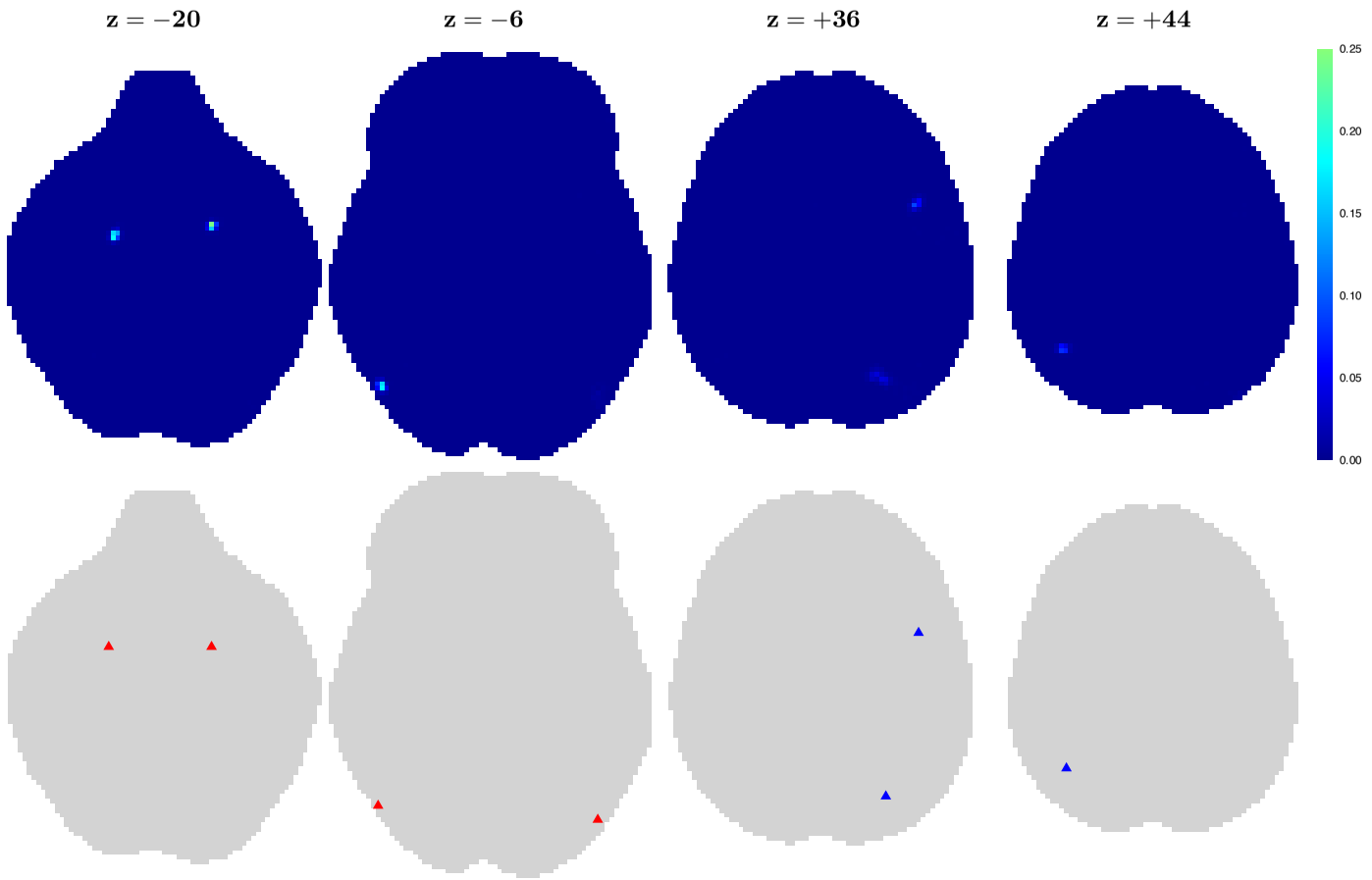



Figure 5.5: Results for the simulated dataset of Section 5.4.2. The figure shows the marginal posterior probability of a focus being at each voxel (top row) and the true locations of study centers, for several axial slices. Red triangles represent the study centers for increases whereas blue triangles represent the study centers for decreases. The probabilities are obtained based on 5,000 posterior samples. The maximum probability observed was roughly 0.247, near the left amygdala.

### 5.4.3 Sensitivity to $m$

A potential weakness of the model is that the background probability parameter $m$ is fixed throughout simulation and not estimated based on the data. Therefore, in this section we examine the effect that changes in the value of $m$ have on our inferences. In Section 5.2 we explain that $m$ is set based on 2 prior parameters, the expected number of background voxels $q$, and the prior beliefs regarding the spread of effects around individual components $\mathbf{S}_{ij}$. Hence, a full sensitivity analysis would require alternating both parameters at the same time and comparing the different outputs. As a first step, we will keep $q$ fixed to 95% and differ the IW $(\mathbf{T}_{\mathbf{x}}, 10)$ prior of $\mathbf{S}_{ij}$.

We generate 25 synthetic datasets based on the simulation setup of Section 5.4.1. For each dataset we try 4 different values for $\mathbf{T}_{\mathbf{x}}$: 1) $\mathbf{T}_{\mathbf{x}} = 40 \times \mathbf{I}_3$; 2) $\mathbf{T}_{\mathbf{x}} = 90 \times \mathbf{I}_3$; 3) $\mathbf{T}_{\mathbf{x}} = 160 \times \mathbf{I}_3$; 4) $\mathbf{T}_{\mathbf{x}} = 250 \times \mathbf{I}_3$. These specifications mean that the mode of the prior on $\mathbf{S}_{ij}$ is isotropic and has a FWHM of roughly 4mm, 6mm, 8mm and 10mm for each dimension, respectively. The values of FWHM are chosen in order to resemble the spatial extent of activations encountered in practice, see for example Xu *et al.* [2009]. Note that $m$ is a decreasing function of the diagonal elements $(\mathbf{T}_{\mathbf{x}})_{ii}$, $i = 1, 2, 3$. Fixed parameters are the same as in Section 5.4.2. In particular, we set $\epsilon = \frac{2}{|\mathcal{B}|}$, $\eta = 2$ and $\mathbf{T}_0 = \frac{5}{30}$. For $\boldsymbol{\lambda}^+$ and $\boldsymbol{\lambda}^-$ we draw uniformly at random from the posterior intensities of emotions and executive control studies, as obtained from the meta-analysis of Section 4.5; for each dataset we use a different posterior draw but we then keep the same draw with all 4 priors on $\mathbf{S}_{ij}$. Again, we run MCMC for 65,000 iterations, discard the first 15,000 as burn-in and thin the remaining samples every 10 iterations to estimate any quantity of interest.

For each dataset and prior we can use the posterior draws $n(\mathbf{x}_i)^{\ell}$ and $n(\mathbf{z})^{\ell}$ to estimate $\mathbb{P}(n(\mathbf{x}_i)) = k$ and $\mathbb{P}(n(\mathbf{z})) = k$, the posterior probabilities of observing exactly $k$ components per individual and exactly $k$ study centers. These are shown in Figures 5.6 and 5.7, respectively. Even though we expect that both of these quantities are estimated as 8 (the true values used in the simulation), we can see in Figure 5.6 that the 4 different priors result into differences in the total number of components per individual. In particular, we see that higher values of $m$ are associated with higher values for $n(\mathbf{x}_i)$. We attribute this behaviour to the fact that as $m$ increases, it becomes harder for voxels that are located relatively far from a component to be assigned to it; as a result we need more components, each one being assigned a small number of voxels around it. Indeed, the median values for the diagonal entries of $\mathbf{S}_{ij}$ for priors 1-4 were 5.32, 8,87, 11.01 and 12.41 respectively. On the contrary, the distribution of the total number of study centers is more robust

to the specification of $m$; in all 4 priors the posteriors are peaked near 8, the true value. However, it is still the case the most conservative value of $m$ assigned by prior 1 resulted into some extra study centers compared to the other priors, see top left panel of Figure 5.7. The explanation is that since for prior 1 there are on average 20 individual components per participant and the expected number of components per foci is $\eta = 2$, there need to be more foci to account for these components. The update of parameter $\eta$ will be therefore considered in our future work.

For a given voxel, one can use the posterior probability of a positive (increase) effect and classify the voxel as activated when this probability exceeds a certain threshold $\alpha$. Analogously, a voxel is classified as demonstrating a negative effect if the posterior probability of deactivation is above $\alpha$. In Figure 5.8 we calculate the true positive rate, the proportion of voxels that were classified as showing an effect (either positive or negative) among the voxels that we truly activated when the data were generated. Results indicate sensitivity of the classification performance to the choice of the prior. When $m$ increases, posterior probabilities are generally lower which in turn results in worse power properties. We therefore suggest that real data analyses are followed by a sensitivity analysis on the background probability in order to make sure that the effect detection is not degraded by $m$. Regarding the false positive rate, the proportion of background voxels that are classified as activated by the algorithm, we found that even with the most tolerant prior, prior 4, on $m$ the rate of false positives did not exceed 0.01% for a very low threshold of $\alpha = 0.1$. This suggests that the method has a good control of false positives. However, such performance is anticipated because BOLD responses for background voxels were drawn independently one of another and so very few positive or negative spikes in the $T$ images appear, except from the locations around which we have place effects. In practice, this assumption may be violated due to preprocessing of the data with spatial normalisation or smoothing. We will consider such simulation scenarios in our future work.

## 5.5 Application

### 5.5.1 Data description

Our application focuses on the Hariri emotional faces task [Hariri *et al.*, 2002]. In this task, subjects are simultaneously presented with three pictures; of these, one picture is the fixed and participants are asked to indicate which one of the other two is identical to the fixed one. Pictures represent faces of humans that have some emotional expression, e.g. fearful or angry. It is widely accepted that

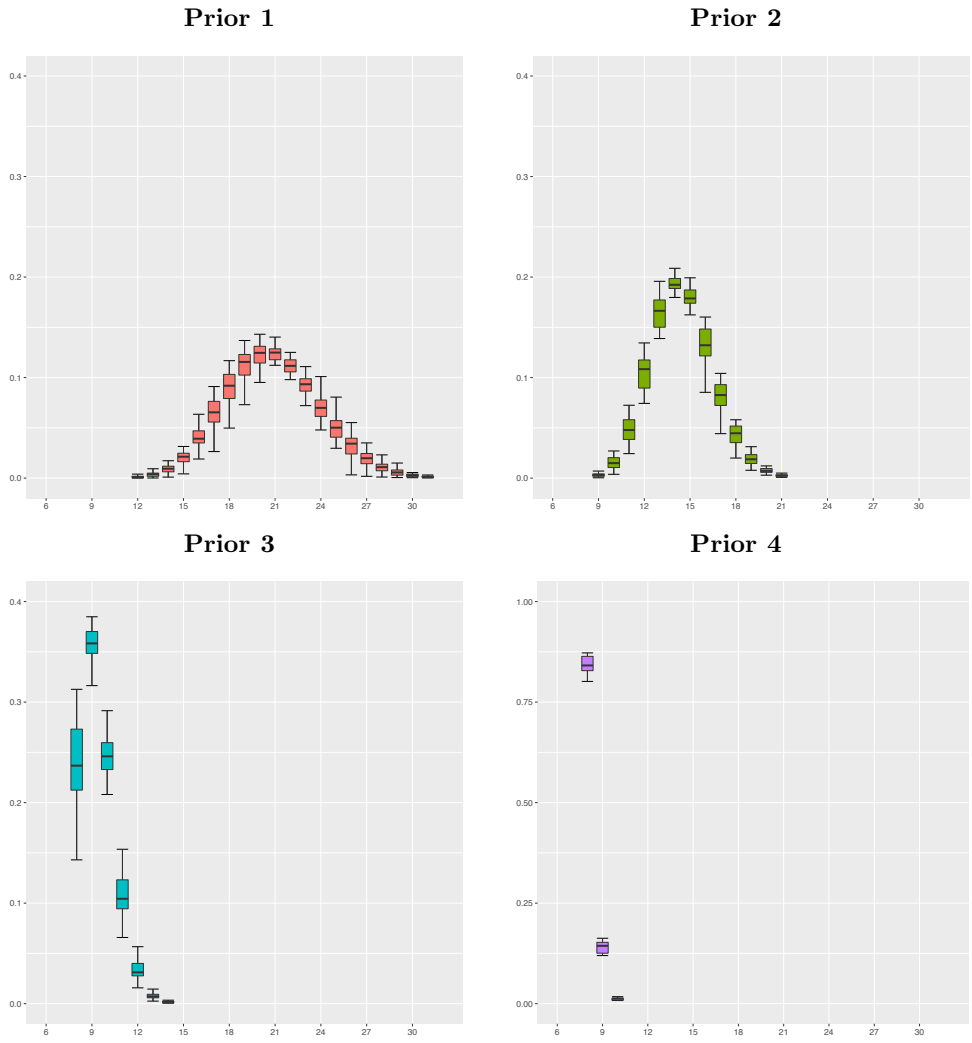



Figure 5.6: Sensitivity analysis results. In this plot we show posterior probabilities of observing exactly $k$ individual components, $k = 6, \ldots, 15$. In each setup, barplots represent the Monte Carlo distribution of the medians (over participants) obtained from the same 25 simulated datasets, generated under the mechanism proposed in Section 5.4.1. For every dataset, we run MCMC for 65,000 iterations, discard the 15,000 first as a burn-in and apply a thinning factor of 10 to the remaining for a total of 5,000 posterior draws.

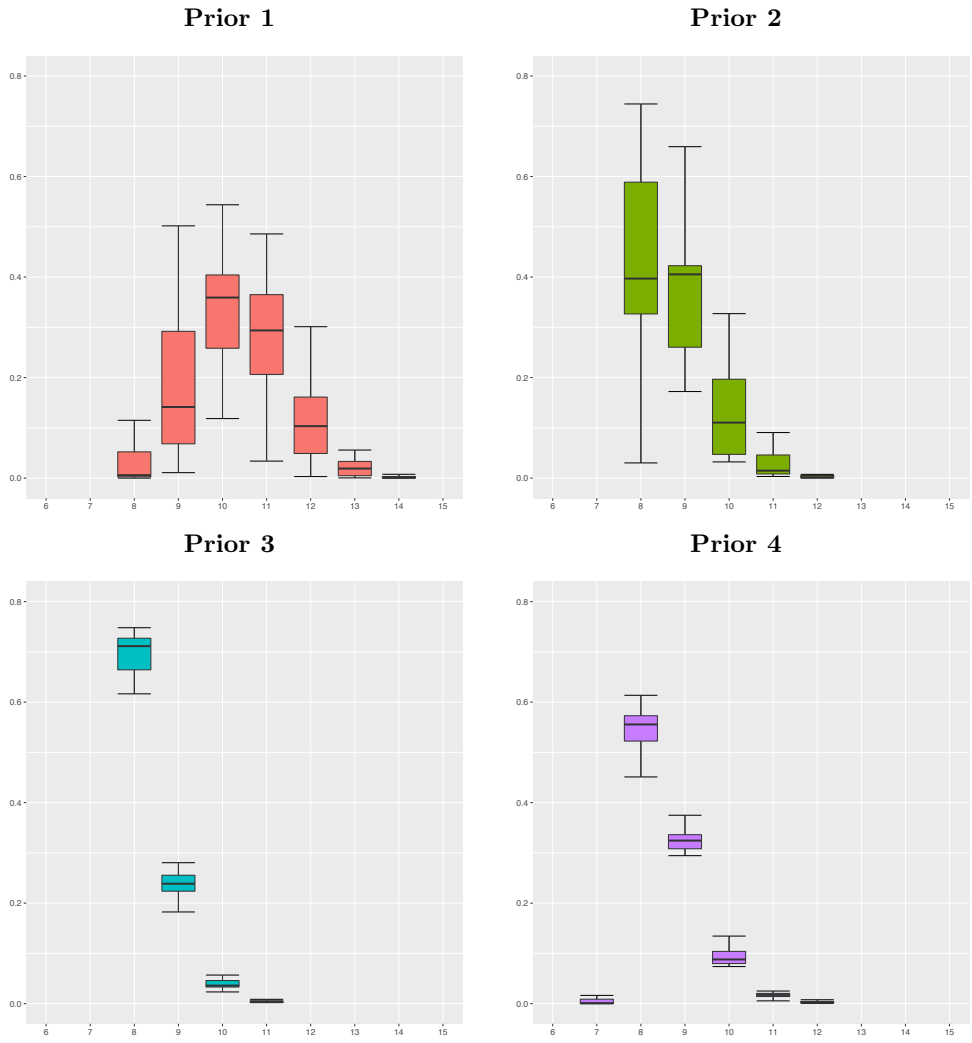



Figure 5.7: Sensitivity analysis results. In this plot we show posterior probabilities of observing exactly $k$ study centers, $k = 6, \ldots, 15$. In each setup, barplots represent the Monte Carlo distribution of the medians (over participants) obtained from the same 25 simulated datasets, generated under the mechanism proposed in Section 5.4.1. For every dataset, we run MCMC for 65,000 iterations, discard the 15,000 first as a burn-in and apply a thinning factor of 10 to the remaining for a total of 5,000 posterior draws.

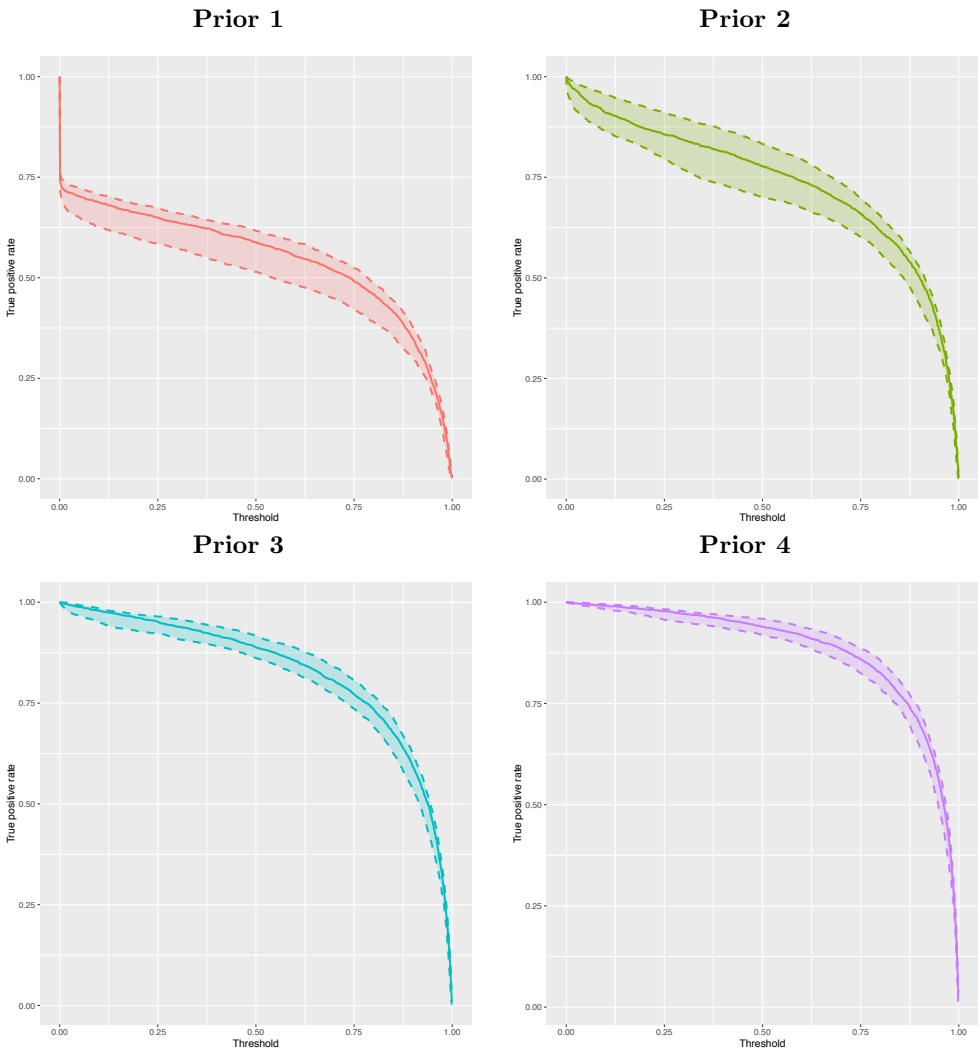



Figure 5.8: Sensitivity analysis results. In this plot we show the true positive rate, the proportion of truly activated voxels over all participants that are assigned posterior probability of showing an effect (increase or decrease in response) that is above a given threshold $\alpha \in [0, 1]$. Solid and dashed lines represent the median and 95% intervals obtained by the the same 25 datasets which we generate under the mechanism proposed in Section 5.4.1. For every dataset, we run MCMC for 65,000 iterations, discard the 15,000 first as a burn-in and apply a thinning factor of 10 to the remaining for a total of 5,000 posterior draws.

such stimuli require the engagement of several regions of the brain, including the amygdalae [Davis and Whalen, 2001] which in Section 4.5 are responsible for emotion processing.

In their study, Hariri *et al.* [2002] tested for the effect of the faces task compared to a baseline condition where the 12 participants were matching simple geometrical shapes instead of faces. They found, using an uncorrected $p < 0.05$ threshold, that both the right and left amygdala were activated by the emotional faces task. Other activated regions included the ventral prefrontal cortex, the posterior fusiform gyri and parahippocampal gyri. However, some of the effects were subtle (e.g. $t$ score for the focus found in the left amygdala was 1.88) and so it is interesting to see if our model can replicate the result using a different group of individuals.

We use data from the *Human Connectome Project* [Van Essen *et al.*, 2012, 2013, HCP]. Launched in 2009, HCP consists of two main consortia, each involving several institutions, with the main objective of the project being to characterise both the function and the connectivity (network behaviour) of the healthy human brain. HCP uses high resolution scanners as well as carefully defined experimental protocols which results in data of very high quality. To date, nearly 1,000 individuals have been scanned while performing a wide range of tasks (including the faces-shapes task) and the information has become publicly available[2]. We therefore use this dataset for our illustration.

The objectives of this group study is to use the information obtained from the meta-analysis of emotion experiments presented in Chapter 4 in order to identify the regions where the emotional faces task induces a significant increase or decrease in BOLD signal. For the purposes of this initial analysis, we will use a small sample size of $I = 6$ randomly selected participants, exactly half of the sample size in the study by Hariri *et al.* [2002], in order to resemble a scenario of an underpowered study.

### 5.5.2 Implementation details

Due to movements of participants during scanning process, the $T$ images contain several missing values. More specifically, for participants 1-6 we have 6,721 (3.1%), 7,691 (3.6%), 6,029 (2.8%), 6374 (2.9%), 7,114 (3.3%), and 6,741 (3.1%) voxels with missing values. The union of these regions includes 12,600 voxels or 5.8% of the brain. Most of the missing values are located near the brain stem.

[2] `http://www.humanconnectome.org/documentation/S900/`

The missing observations are imputed with the following approach. For each participant we take the data that fall between the 0.05% and 99.95% percentiles of the empirical distribution because we believe that they belong to the background. Based on these, we estimate the empirical mean $\hat{y}$ and variance $\hat{\sigma}_y^2$ which we then use to assign a $\mathcal{N}\left(\hat{y}, \hat{\sigma}_y^2\right)$ value to all voxels with missing data, independently one of another.

We choose not to limit the analysis to a smaller atlas, the intersection of subjects' brains that contain no missing values in order not to prevent effects that are common only to a subset of the participants from showing. Nevertheless, the analyses revealed no interesting study effects in these regions. In Figure B.7 of Appendix B.3 we show how many participants per voxel had missing values, for several axial slices of the brain.

We use the following prior specifications. The data have been smoothed with an isotropic 4mm FWHM Gaussian kernel and so we set the probability that a voxel belongs to the background at $q = 99\%$, where a high probability is chosen in order to prevent false positives. The anticipated task effects are isotropic with a FWHM of 8mm (4 voxels) in each dimension; this is achieved by letting $\mathbf{S}_{ij} \sim \text{IW}\left(\frac{14\times 8^2}{8\log 2}, 10\right)$. The value of $m$ that we obtain based on $q$ and $\mathbf{T}_{\mathbf{x}}$ is roughly 0.16. Since we have a stringent $m$, we expect that each activation will require several components and so we let $\eta = 4$. In order to allow for an average of 2 individual components that do not cluster around any study center we use $\epsilon = \frac{2}{|\mathcal{B}|}$. We expect higher between-subject variability compared to the within-subject variability that is represented by $\mathbf{T}_{\mathbf{x}}$, and hence set $\mathbf{T}_0 = \frac{1}{9}\mathbf{I}_3$; we then *a priori* have that $\mathbb{E}\left[\mathbf{\Sigma}_z\right] = 15$.

For increases, prior information comes from the meta-analysis of 855 studies of emotion that was carried out in Section 4.5. At each iteration of the MCMC we draw $\boldsymbol{\lambda}^+$ uniformly at random from the 1,000 available draws of $\pi\left(\boldsymbol{\lambda} \mid \{\mathbf{u}_i\}_{i=1}^{855}\right)$, where $\mathbf{u}_i$ are the foci of study $i = 1, \ldots, 855$. For decreases we have no prior information and so we use a fixed homogenous $\lambda^- = \frac{4}{|\mathcal{B}|}$ which suggests that we expect 4 study decrease centers.

Posterior samples are obtained using the hybrid Gibbs algorithm developed in Section B.2. We run the algorithm for 100,000 iterations and discard the first 20,000 as a burn-in. All results are obtained after a thinning factor of 16 is applied and so we have 5,000 samples. Convergence is assessed by visual inspection of posterior traceplots for the scalar parameters of the model. Those can be found in Appendix B.3. Note that we run 2 chains with different initial conditions and compare the outputs. The traceplots indicate that the chains have both reached their common stationary distribution and so we conclude that the algorithm has converged.

### 5.5.3 Results

The traceplot for $n(\mathbf{z}^-)$ can be found in Figure B.3. We see that the total number of decrease study centers in the posterior is 2 with probability 65%. Nevertheless, these centers do not cluster around any constant location and hence cannot be associated with any effect. We therefore conclude that there is not enough evidence for the existence of a negative effect of the faces task. This confirms the results of [Hariri *et al.*, 2002].

For increases, there are several areas that show activation. The region around which the strongest effects are present is the fusiform face area (bilateral activation in top left and top middle panel, Figure 5.10). Our finding is an agreement with Hariri *et al.* [2002] and confirms a well-known result that relates the fusiform face area with the perception of faces, see e.g. Kanwisher and Yovel [2006]. Other regions activated by the task are the occipital fusiform gyrus (Figure 5.10, top right panel), the occipital pole (Figure 5.10, bottom left panel), the intracalcarine cortex (Figure 5.10, bottom middle panel) and the inferior frontal gyrus (Figure 5.10, bottom right panel). No activations are found in the amygdala.

In light of our findings in the simulation study of Section 5.4, we perform additional analyses to examine the effect of prior parameters $q$ and $\eta$. In particular, we run another 2 MCMC chains (chains 2-3) where all except one prior parameters are kept fixed at each time: in the second run, we assume that the expected proportion of $q$ of background voxels is 95% instead of 99% and in the third run we increase the expected number of components per center $\eta$ from 4 to 6. Marginal posterior probabilities of observing an increase study center at voxel $v$ are shown in Figures B.5 and B.6 respectively. We see that the spatial location of the study centers are in close agreement between the 3 runs. However, we find that there is an effect of the prior to the total number of study centers as can be seen in Figure 5.9. As expected, increasing $\eta$ has led to a drop of $n(\mathbf{z}^+)$ because more components can be assigned to the same center. Finally, we also observe that reducing $q$ leads to roughly 4 more study centers. The finding is surprising given we would expect that a more lenient background probability to reduce the total number of centers and hence requires further investigation.

To conclude, in Figure 5.11 we show posterior probabilities of activation and deactivation for participant 1 of the study, in axial slices around which study effects were prevalent. We see that the algorithm detects the regions where the highest activations occur and places individual components in mainly in those regions (top row, black dots). Results for participants 2-6 can be found in Figures B.8-B.12 of the Appendix. We note that even though the effects of the participants spread across

several voxels, the model manages to achieve good spatial precision. Regarding the effects in the amygdala, we see that only participant 6 appears to be activated near that region (see top left and top middle panel of Figure B.12) and the remaining participants only show moderate values.

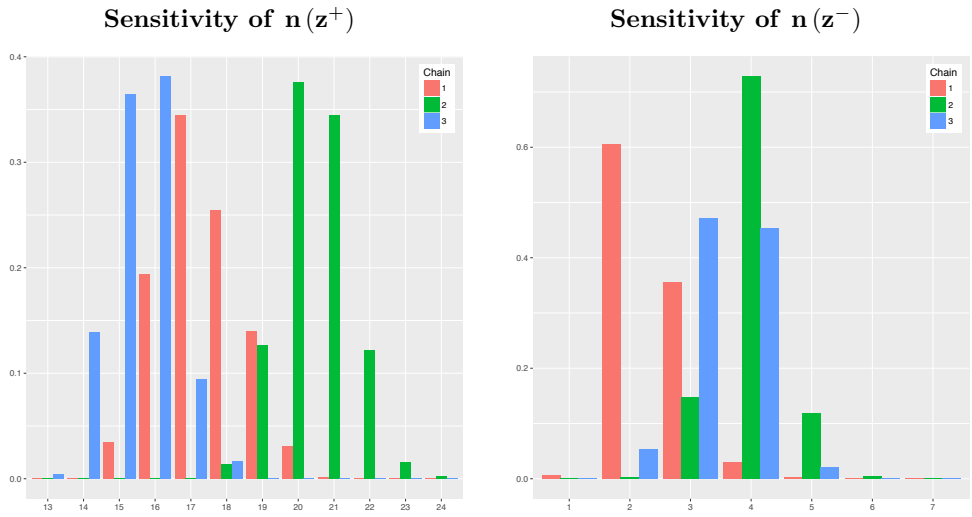



Figure 5.9: Real data sensitivity analysis results. Here we plot the probability of observing exactly $k$ increase (left) or decrease (right) study centers. Run 1 (red) corresponds to our initial run with prior specifications defined in Section 5.5.2. In run 2 (green) we decrease the expected proportion of background voxels to 95% compared to the first run, keeping all other parameters fixed. In run 3 (blue) we use the same prior specification as in run 1 except that we increase the expected number of offsping $\eta$ to 6. Results reveal some sensitivity to the choice of the prior.

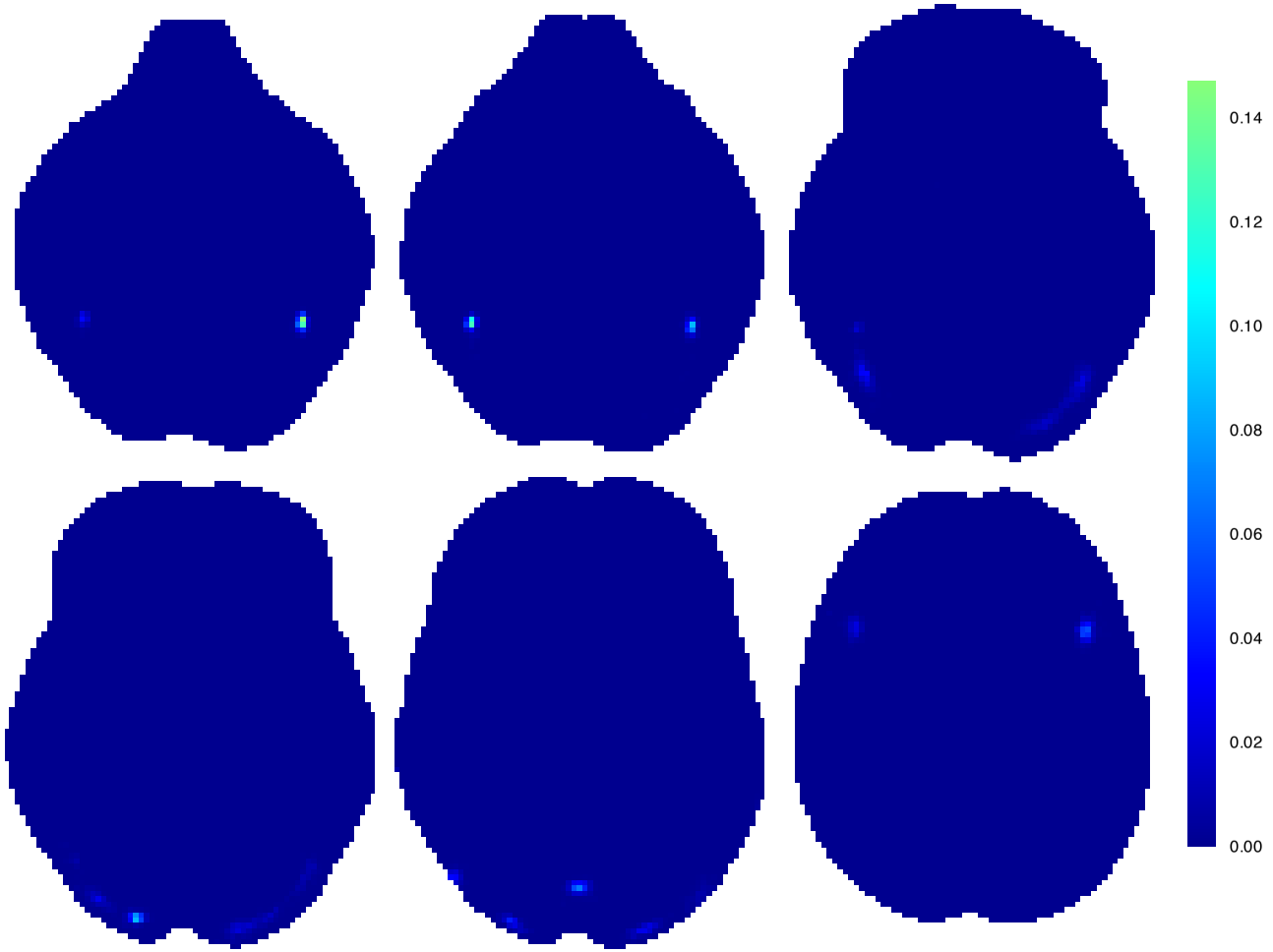



Figure 5.10: Group fMRI study results. The plot shows the marginal posterior probabilities of observing a study center in voxel $v$. The activated regions are: fusiform cortex (top left and middle panel), occipital fusiform gyrus (top right panel), occipital pole (bottom left panel), intracalcarine cortex (bottom middle panel) and inferior frontal gyrus (bottom right panel). The probabilities are based on 5,000 samples from the posterior.

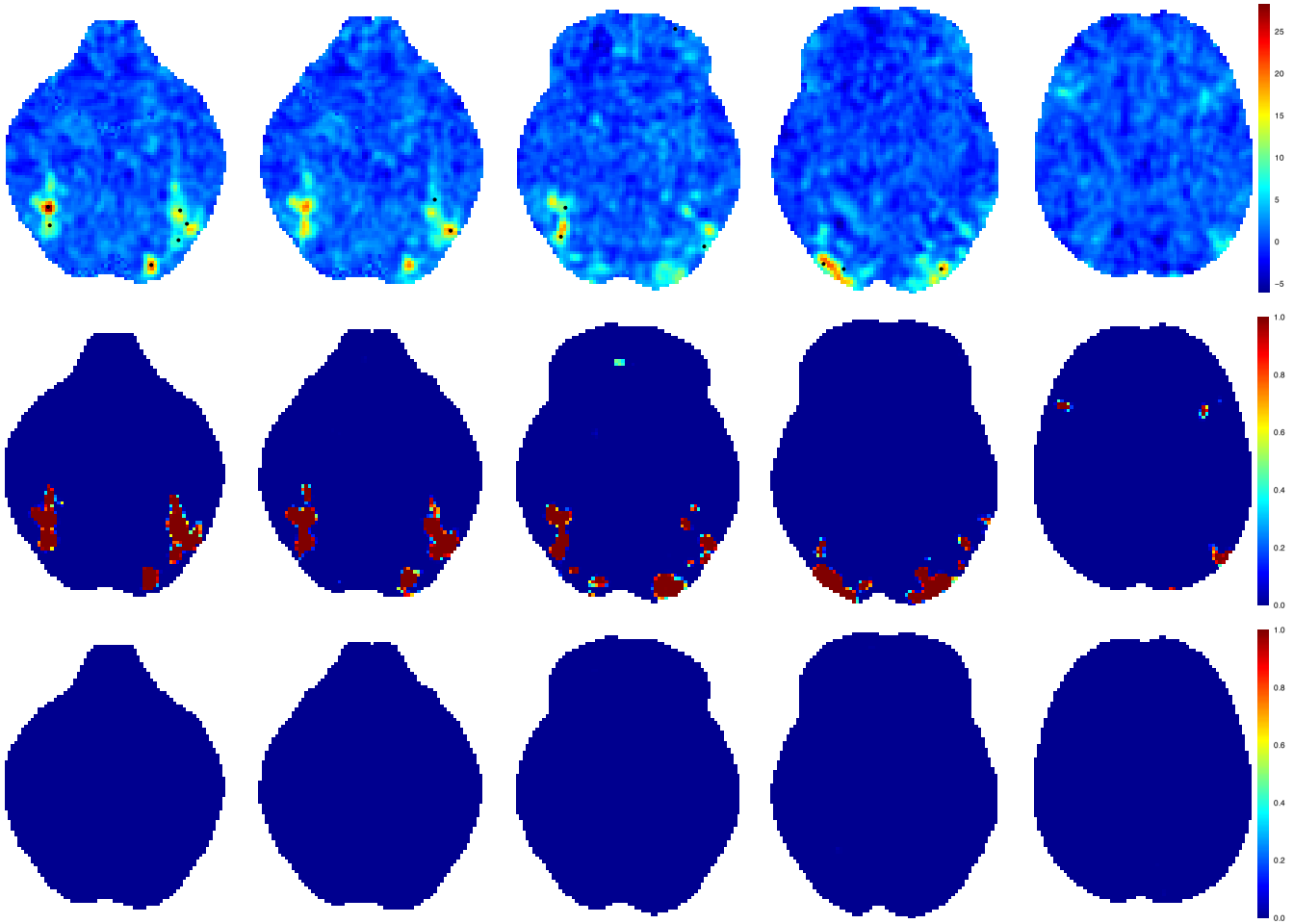



Figure 5.11: Group fMRI study results for participant 1. The figure shows the observed data (top row), the voxel-wise probability of increase (middle row) and the voxel-wise probability of decrease (bottom row). Estimates are obtained based on 5,000 posterior draws. The black dots represent the individual center process $\mathbf{x}_1$ at iteration 100,000 of the MCMC. The columns correspond (from left to right) to axial slices $z = -20$, $z = -18$, $z = -14$, $z = -8$ and $z = +26$.

## 5.6 Discussion

In this chapter, we have developed a Bayesian hierarchical model for group fMRI studies. The model extends the work of Xu *et al.* [2009] by accounting for negative BOLD responses and using CBMA data as prior information in the analysis. Further, our model builds upon the spatial point process framework, thus facilitating the use of the spatial birth-and-death algorithm [Preston, 1977] which demonstrates better convergence and mixing properties compared to the reversible jump MCMC algorithm that was used by Xu *et al.* [2009].

Some initial results from simulation studies suggest that the model performs well in capturing both increases and decreases in BOLD response, even when the observed signal is low. Application on a study of the emotion faces task [Hariri *et al.*, 2002] replicated some of the results of previous studies with an extremely small sample size, while also attaining good spatial precision. Even though this is partly due to the high quality of the data in hand, it also demonstrates the benefits of spatial models compared to the standard massively univariate approaches.

Nevertheless, our model was found to be sensitive to the choice of some prior parameters, especially the prior on the background probability $m$. In our future work we will try to extend the model so that $m$ is estimated along with the remaining model parameters. Another direction that we will take is to explain through a simulation study the setups in which the use of CBMA prior information benefits the most compared to the homogenous prior. This characterisation is meaningful both in terms of the maximum effect size but also the maximum number of participants for which our approach outperforms the standard approach.

A potential drawback of our approach is that it requires a significant amount of computational effort compared to the massively univariate approaches. Even though application of our model using multiple processors with OpenMP reduces the total running time significantly, it still took roughly 65 hours to run 100,000 iterations on the real data with an Intel i7 3.4GHz with 4 cores. A possible solution to this problem would be to carry out the voxel-wise calculations of the lower level using a GPU. We will consider this alternative in our future work.

The study of the faces task with data from the HCP will be continued and focus on understanding its effect on the amygdalae. Even though failure to replicate previous findings could be because of our small sample size, it could also be the case that our birth-and-death proposals generate components mainly in the region of the gyri where the strong effects appear and so it is not possible for centers to be generated near the amygdala. Using a homogenous proposal would cancel

the benefits of the current approach but adopting a mixture proposal where with probability $1 - \alpha$ we use the current proposal and with probability $\alpha$ we use a proposal based on the CBMA prior is a promising option the we aim to explore.

Finally, another direction that we will consider is to include image-based meta-analysis (IBMA) data as an input in our model. For each voxel, we will assume that the corresponding IBMA value arises from a mixture where the mixing components will represent either background noise or some study center. Inference can be achieved with a spatial birth-and-death algorithm, similar to the one that we used to simulate the individual component process in our model.

# Chapter 6

# Estimating the number of missing studies in neuroimaging meta-analysis

## 6.1 Introduction

The first step in a meta-analysis is a literature search [Normand, 1999]. During this step investigators use databases to retrieve all previous work which is relevant to the question of interest. Ideally, this process will yield an exhaustive or at least representative sample of studies on a specific topic. Unfortunately, literature search is subject to the *file drawer* problem [Rosenthal, 1979; Iyengar and Greenhouse, 1988]. File drawer refers to research studies that are initiated but are not published due to lack of significance, either by cause of authors' hesitation to submit or perhaps because of rejection by journals that are reluctant to publish negative results. The file drawer along with the other forms of publication bias (see Song *et al.* [2000] for an overview) can potentially undermine the quality of a meta-analysis as they lead to biased estimates of the effect of interest [Hedges and Olkin, 1985; Begg and Berlin, 1988; Hedges, 1989; Sutton *et al.*, 2000a]. Aside from distorting a particular scientific question of interest, this feeds into researchers' scepticism regarding the usefulness of meta-analysis [Greenland, 1994].

Evidence of the file drawer problem has been found in many fields of scientific research including medical research [Begg and Berlin, 1989; Easterbrook *et al.*, 1991; Dwan *et al.*, 2008, 2013] and the social sciences [Sterling *et al.*, 1995]. Several methods have been proposed for detecting and sometimes adjusting for the presence of the file drawer problem. Early literature on the topic was focused on finding the

*fail-safe N* [Rosenthal, 1979; Iyengar and Greenhouse, 1988], the minimum number of unpublished studies required to overturn the outcome of meta-analysis. Much attention has been given to the graphical tool known as the *funnel plot* [Light and Pillemar, 1984]. The funnel plot is based on the assumption that point estimates from large studies should be more tightly scattered about the true effect than point estimates from smaller ones; and in all instances they should be symmetrically distributed about the true parameter value. Thus a plot of effect precision (e.g. sample size) vs effect estimate should produce a funnel shape; any asymmetries or trends in the funnel plot are indications of publication bias. Many tests and methods have been developed to formalise the idea of funnel plot including Begg and Mazumdar [1994], Egger *et al.* [1997], Duval and Tweedie [2000a] and Duval and Tweedie [2000b].

Another very common approach involves the use of *weight functions* where the probability of observing a data point is modelled as a function of its characteristics, e.g. $p$-values. Several weighting schemes have been proposed, see for example Hedges and Olkin [1985], Iyengar and Greenhouse [1988], Hedges [1992] or Larose and Dey [1998] for implementation under a Bayesian perspective. Copas and Jackson [2004] give an upper bound for the total number of missing studies under any possible weight function. Finally, another popular approach is *sensitivity analysis* where the probability of publication is given a parametric form and the outcome of meta-analysis is studied under different values of the parameters [Copas, 1999; Copas and Shi, 2000; Copas, 2013]. For a better overview and more detailed description of methods for modelling the file drawer problem we refer the reader to Sutton *et al.* [2000b], Rothstein *et al.* [2006] and Jin *et al.* [2015].

In the field of fMRI, there has been little study of biases in meta-analysis. One example is David *et al.* [2013], who studied the relation between sample size and the total number of activations to conclude the existence of reporting biases, mainly affecting small studies. However, to date there had been no work on estimating the fundamental file drawer quantity, the number of missing studies. In what follows, we propose a model for estimating the number of non-significant studies omitted from a large cohort of studies in the context of CBMA and discuss possible implications. The data that we use are presented in the following Section 6.2. The details of the model are presented in Section 6.3, where we also summarise our findings. In Section 6.4 we repeat the meta-analysis of emotions studies of Chapter 4 after imputing the estimated file drawer studies and compare the results. Finally, in Section 6.5 we conclude with an evaluation of our model and some possible extensions.

## 6.2 The BrainMap database

Our analysis is motivated by coordinate data from *BrainMap* [Laird *et al.*, 2005]. BrainMap is an online, freely accessible database for coordinate-based data of both functional and structural neuroimaging experiments. The database is continuously expanding and up to date consists of results obtained from 2,562 scientific papers, each one containing several experiments or, *contrasts*. BrainMap is an invaluable resource for meta-analyses and indeed many studies have been based on data retrieved from the database (see Hill *et al.* [2014] and Kirby and Robinson [2015] for some recent examples). It is therefore of vital importance to investigate the presence of the file drawer problem and its possible effects on a meta-analysis.

From here on, we follow the convention of considering our unit of observation contrasts (instead of studies), and hence our dataset consists of 12,292 observations. Each data point consists of a set of three dimensional coordinates $\mathbf{x}_i$, the *foci*, which correspond to the centers of the clusters where significant activation occurred during the experiment. For the purposes of this work we suppress the spatial aspect of the problem and instead model the file drawer only based on the raw foci counts $n_i$, the total number of foci per contrast. Figure 6.1 shows the empirical probability mass function of this count data.

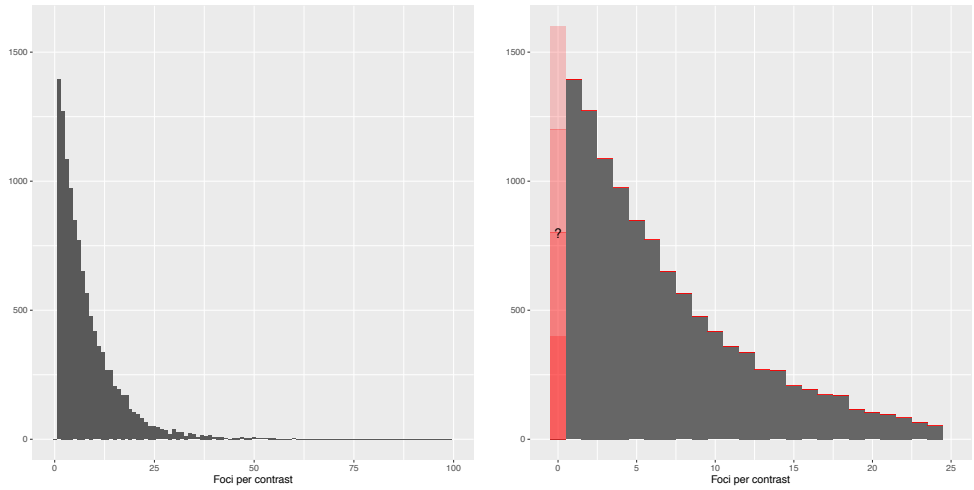



Figure 6.1: Empirical distribution of the total number of foci per experiment in the BrainMap database, $n_i$. Left: full distribution. Right: zoom in. No instances of studies which report zero counts occur.

The barplot of Figure 6.1 (right) identifies a striking aspect of this data: even though the distribution of $n_i$ has most of its mass close to zero, there are

no instances of studies which report zero foci. Of course, such instances exactly correspond to the experiments with no significant findings and so missingness can be attributed to publication bias. The severity of the phenomenon relates to the unobserved data that is the studies for which $n_i = 0$ and are suppressed from the literature. In next section we attempt to estimate the total number of zero-count studies that are missing from the BrainMap database.

## 6.3 A zero-truncated count model for CBMA file drawer

Our model uses the count data from the observed, significant studies to infer on the file drawer quantity. At this point we make some critical assumptions: I) the full data $\{n_i\}_{i=1}^{I}$, observed and unobserved, are independent and identically (i.i.d.) samples from a count distribution $N$ with a certain parametric form; II) the probability of publication depends on the total number of significant activations $n_i$, specifically there is probability zero for studies with $n_i = 0$ and probability one for studies with $n_i \geq 1$. To ensure that the assumption of independence is valid we use 5 subsamples of the data (A-E), where in each subsample only one contrast from each publication is used. If possible we use 5 different contrasts, one in each of the 5 datasets; for publications with less than 5 contrasts we ensure that every contrast is used at least once and then randomly select one for the remaining subsamples. For a detailed discussion of implications of assumptions I-II see Section 6.5.

If assumptions I-II described above hold, then a suitable model for the data is a *zero-truncated* count distribution. A zero-truncated count distribution occurs when we restrict the domain of a count distribution $\pi(n \mid \theta)$ to the positive integers:

$$\pi_{\mathrm{ZT}}(n \mid \boldsymbol{\theta}) = \mathbb{P}(N = n) = \frac{\pi(n \mid \boldsymbol{\theta})}{1 - \pi(0 \mid \boldsymbol{\theta})}, \quad n = 1, 2, \ldots, \tag{6.1}$$

where $\boldsymbol{\theta}$ is the parameter vector. The families of count distributions that we consider are the Poisson and Negative Binomial. For the Negative Binomial distribution we use the mean-dispersion parametrisation:

$$\pi(n \mid \mu, \phi) = \left(\frac{\phi}{\phi + \mu}\right)^{\phi} \frac{\Gamma(\phi + n)}{\Gamma(\phi)} \left(\frac{\mu}{\mu + \phi}\right)^{n}, \tag{6.2}$$

where $\mu$ is the mean, $\phi$ is the dispersion parameter and $\Gamma(\cdot)$ represents the gamma function.

Once the parameters of the truncated distribution are estimated one can make statements about the original, untruncated distribution. The key quantity for

any inference is the probability of observing a zero count study $\pi(0 \mid \theta)$. One possible way to express the file drawer quantity is the prevalence of zero count studies $p_z$, the total number of missing experiments per 100 published which can be estimated as:

$$\hat{p}_z = \frac{\pi(0 \mid \hat{\theta})}{1 - \pi(0 \mid \hat{\theta})} \times 100. \tag{6.3}$$

To explain as much nuisance variability as possible, we further model the expected number of foci per experiment as a function of its characteristics in a regression:

$$\mathbb{E}\left[N_i \mid \mathbf{x}_i\right] = \exp\left(\mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}\right), \tag{6.4}$$

where $\mathbf{x}_i$ is the vector of covariates and $\boldsymbol{\beta}$ is the vector of regression coefficients. The covariates considered include: i) the year of publication, ranging from 1985 to 2014 with median 2004; ii) the (square root) number of participants ranging from 1 to 395 with median 12; iii) the context of the experiment with levels age effects, disease effects, drug effects, gender effects, learning, linguistic effects, normal mapping and other (any other label that appears less than 20 times).

Parameter estimation is done under the Bayesian paradigm using MCMC. We use a simple random walk Metropolis step with normal proposal to update the vector $\boldsymbol{\theta}$. The covariance of the proposal is tuned during burn-in using the adaptive algorithm of Haario *et al.* [1999], and is then kept fixed throughout simulation. All results are based on MCMC runs of 100,000 iterations thinned every 100 iterations, after a burn-in period of 20,000 iterations. The properties of our sampler for the simple zero-truncated Negative Binomial model are studied on synthetic data through a simulation study. The results are presented in Appendix C.3. For completeness, we compare our estimates with the estimates obtained under the *Generalized Additive Models for Location Scale and Shape* (GAMLSS) framework of Rigby and Stasinopoulos [2005]. Fitting is done in R [R Core Team, 2015] with the *gamlss* library [Stasinopoulos and Rigby, 2007]. We now summarise our findings.

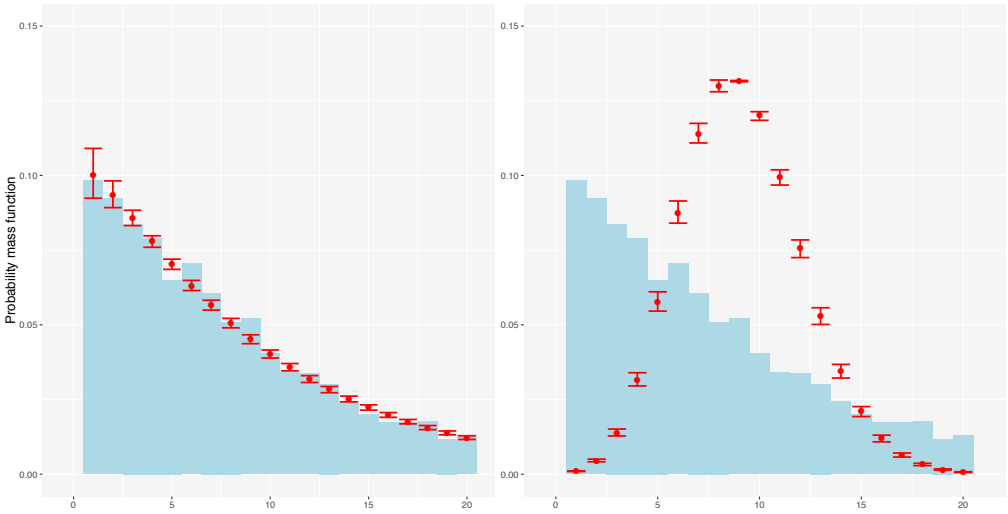
With no covariates the fit is easily visualised. Figure 6.2 shows results for subsample A; results for the remaining subsamples can be found in Figure C.1 of the Appendix. We see that the Negative Binomial distribution provides a much better fit compared to the Poisson distribution, due to the overdisperssion that cannot be captured by the Poisson model. This is verified by our posterior predictive checks which we conduct as follows. For both models we approximate the posterior predictive distributions by simulating 1,000 count samples for each one of the posterior draws obtained with MCMC. We then calculate the 95% posterior predictive intervals and record the percentage of real data observations that fall within these

intervals. For datasets A-E, the 95% posterior predictive credible intervals of the zero-truncated Poisson model contain 56.05%, 56.44%, 58.90%, 56.75% and 56.26% of the observations, respectively. The corresponding figures for the zero-truncated Negative Binomial model are 97.26%, 97.53%, 97.38%, 97.25% and 97.69%. We therefore use the zero-truncated Negative Binomial model to obtain estimates of the file drawer quantity, see Table 6.1. The median posterior values obtained from the MCMC are in close agreement with the GAMLSS method, both suggesting the the prevalence of zero-count experiments is roughly 9 per 100 published experiments.

| **Subsample** | **Median** | $\mathbf{p_{0.025}}$ | $\mathbf{p_{0.975}}$ | **GAMLSS** |
|---|---|---|---|---|
| A | 10.11 | 8.68 | 11.93 | 10.14 |
| B | 9.45 | 8.13 | 11.14 | 9.47 |
| C | 8.98 | 7.72 | 10.52 | 9.02 |
| D | 8.71 | 7.42 | 10.11 | 8.67 |
| E | 10.17 | 8.70 | 11.99 | 10.16 |

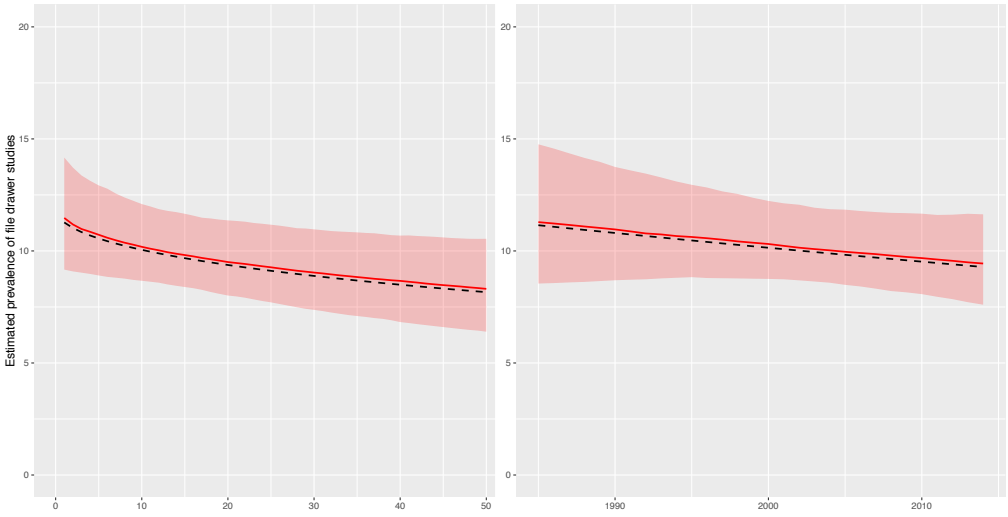
Table 6.1: Estimated posterior prevalence of zero-count studies, as obtained from the 5 subsamples. Columns 2-4 show posterior summaries of the MCMC samples while column 5 is the GAMLSS estimate.

We generally observe that covariates essentially have no effect on the estimated prevalence of missing studies. Due the lack of fit by the zero-truncated Poisson model, we only show results obtained with Negative Binomial regression. The estimated prevalence of zero count studies is a slowly decreasing function of both the number of participants and the year of publication (Fig. 6.3 for subsample A, Fig. C.2 for subsets B-E). For the former, the trend is expected. One possible explanation is that bigger samples result into greater power and therefore more foci. Nevertheless, a stronger effect was anticipated. For the latter, one would expect that the establishing of multiple testing correction techniques through the years would decrease the rate of false positives and hence shift the distribution of foci counts towards zero. Hence, both findings require further investigation. Finally, we see that the estimated prevalence of zero count studies is similar for all levels of the categorical variable context, with the exception of experiments studying gender effects (Figures 6.4 and C.3 for subsamples A and B-E respectively). The median posterior prevalence, setting year and sample size to the median values found in the subsamples and taking the weighted average over the levels of context, is found to be 10.03, 9.20, 8.55, 8.52 and 10.05 for subsamples A-E, respectively.




Negative Binomial Poisson

Figure 6.2: Zero-truncated fit with the Negative Binomial (left panel) and Poisson (right panel) distributions for subsample A. Points represent posterior medians obtained from 30,000 MCMC iterations, thinned every 30 iterations. Error bars represent the 95% credible intervals.




Sample size Year of publication

Figure 6.3: Prevalence of file drawer studies as a function of study sample sample size (left) and study year of publication (right) for subsample A. Red solid line is the posterior median whereas the shaded area indicates the 95% posterior credible intervals. The dashed line represents the estimate obtained with GAMLSS. All values are averages across the levels of study context.

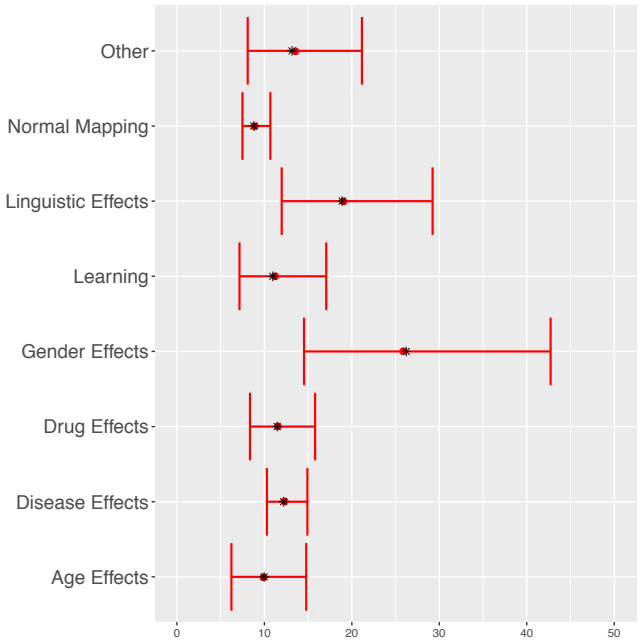



Figure 6.4: Prevalence of file drawer studies as a function of study context for subsample A. Red dots are the posterior medians whereas the error bars represent the 95% credible intervals. Black asterisks are the GAMLSS estimates. Publication year and sample size have been set to their median values.

## 6.4 Effect of missing studies

The zero-truncated regression analysis of the previous sections suggests that if the assumptions of our file drawer model hold and the BrainMap database is representative of the population of CBMA studies, then there are roughly 9 unpublished experiments per 100 that we observe. Therefore, it is plausible to ask what would be the effect of these unobserved, zero count studies on the outcome of a meta-analysis. In this section we repeat the meta-analysis of emotion studies shown in Chapter 4, adding some studies with no foci and demonstrate the effects on the CBMA results that the inclusion of these studies has.

Using our estimates from Section 6.3 we add 9 zero count studies for every 100 of the original dataset which consists of 855 studies. Hence, there are 941 studies in the new dataset of which 86 report no activations. Analysis is carried out with the LGCP model presented in Chapter 4 using no covariate information. HMC is run for 15,000 iterations, with the first 5,000 being the burn-in period. The last 10,000 are thinned every 10 iterations to obtain a posterior sample of 1,000 intensities. These are used to calculate several quantities of interest which we now present.

The median posterior intensities for several axial slices of the brain are shown in Figure C.4. Qualitatively, the results are almost identical to the ones obtained when analysing the original dataset (see Figure 4.6). Nevertheless, there are still differences between the two analyses. For example, the posterior expected number of foci in the original dataset is 7.15 (95% CI [6.99,7.37]); this is significantly higher compared to the estimate obtained when zero count studies are included which is 6.50 (95% CI [6.33,6.67]). In Figure 6.5 we show posterior distributions for the probability of observing at least one focus in the right and left amygdala, the regions which are mainly activated by emotion processing. Even though there is considerable overlap between the two, we see that the posterior median is roughly 10% lower when we add the zero count studies for both left and right amygdala.

Overall, we see that inclusion of zero count studies in the meta-analysis has lead to lower estimates of the effect of emotion processing in brain activation, as we expected. Obviously, the extent to which the file drawer problem inflates the estimates obtained from a meta-analysis depends on the unknown total number of missing studies. It is therefore essential to account for the possibility of missing studies in any CBMA and examine estimates obtained under different scenarios.

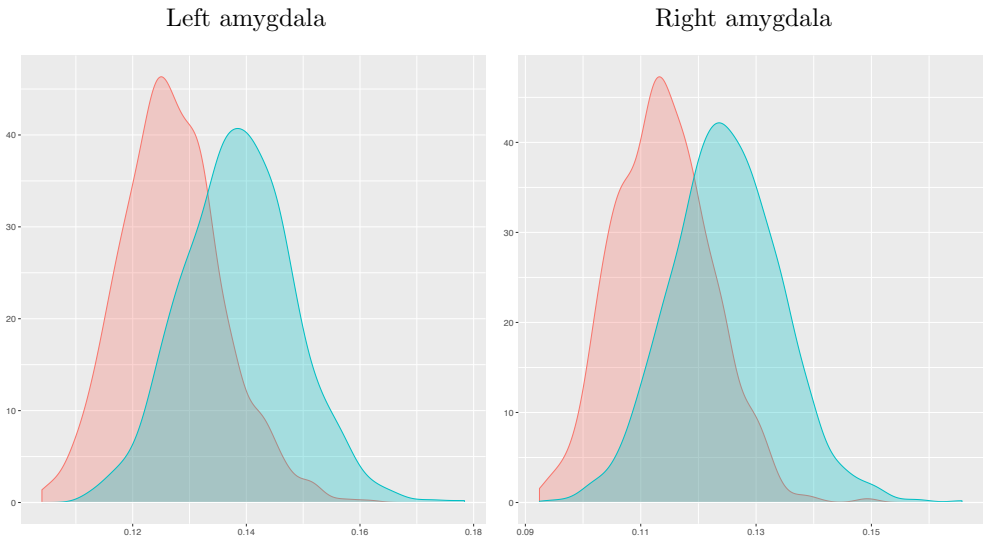



Figure 6.5: Posterior probabilities of activation for the amygdala, as obtained from the original (green) and the new dataset with added zero foci studies (red). The left panel shows results for left amygdala and right panel for the right amygdala. All distributions are calculated based on a sample of 1,000 posterior draws.

## 6.5 Discussion

We have proposed a simple method for estimating the number of missing studies from a meta-analysis. Our method uses intrinsic statistical characteristics of the non-zero count data to infer zero counts. We find that the zero-truncated negative binomial distribution provides a good fit for the total number of foci per contrast in the BrainMap database. The analysis suggests that the magnitude of the file drawer, zero-foci experiments slightly varies depending on study characteristics but is generally around 9 per 100 published experiments. The number is significantly greater than zero which indicates the existence of publication bias in coordinate-based meta-analysis. Some of the missing experiments can be attributed to negative contrasts reported in the original publications but not registered in the database but surely some are never published.

Our approach relies on assumptions I and II described in Section 6.3. Assumption I implies that there is independence between each contrast. However, as one publication can have several contrasts, this assumption is tenuous despite it being a standard assumption for most CBMA methods. To ensure the independence assumption is valid, we subsample the data so that only one randomly selected contrast per publication is used. Assumption II defines our censoring mechanism, such

that experiments with at least one significant activation are always published. The assumption that no significant research findings are suppressed from the literature has been adopted by authors in classical meta-analysis [Eberly and Casella, 1999] and we believe that is reasonable in the context of CBMA as well. For the unobserved studies, we accept that all experiments reporting null results never appear on the database. Note, that since studies typically examine several contrasts of interest, authors have fewer incentives to report non significant experiments because of the other, significant results that can be reported in the publication. At this point, one possible source of bias introduced due to the assumptions of the model is the existence of studies with several non significant experiments; however, in such a case our estimates are only underestimating the file drawer quantity and hence we choose to make no corrections.

Our file drawer model has some limitations. Firstly, the analysis in based on data retrieved from a single database. As a consequence, results are not robust to possible biases in the way publications are included in this particular database. A more thorough analysis would require consideration of other databases and papers that haven't been registered in BrainMap. Secondly, one may argue that our censoring mechanism is rather simplistic. For example we have not allowed for the possibility of experiments initially having negative results but then changing the analysis pipeline (e.g. random vs fixed effects) to finally obtain some significant activations. This would be an instance of initially-censored (zero-count) data being 'promoted' to a non-zero count through some means. Such models can be fit under the Bayesian paradigm and will be examined in our future work.

# Chapter 7

# Conclusions

## 7.1 Contributions

In this thesis, we have considered various aspects of coordinate-based meta-analysis of fMRI data. CBMA is a topic of growing interest in statistics but presents researchers with various challenges. Our main objective has been to develop spatial models that address the limitations of the widely used massively univariate approaches.

In Chapter 3 we review the literature on CBMA, showing the benefits of using the spatial model-based methods compared to the massively univariate kernel-based approaches. Of particular interest are the results of a simulation study regarding the power properties of the activation likelihood estimation algorithm [Eickhoff *et al.*, 2012]. This simulation study demonstrates the fixed effects nature of the algorithm that allows a small proportion of studies to drive statistical inferences. This important issue may well extend to other kernel-based methods. We further identified several still-open problems in the field such as the need for spatial methodologies that can account for co-activations patterns in CBMA data.

In Chapter 4 we attempt to address one of the crucial open problems in CBMA, namely meta-regression, by proposing a novel model based on log-Gaussian Cox processes [Møller *et al.*, 1998]. The results of our simulation studies and experimental data analysis suggest that the model works well even with challenging datasets of sparse point realisations and can be used to compare multi-type point patterns. Additionally, the algorithm that is used for posterior simulations shows very good mixing properties and its running time is greatly reduced by implementation on a GPU leaving room for further improvement as technology on GPUs improves.

Motivated by the growing body of literature in CBMA, in Chapter 5, we have extended the work of Xu *et al.* [2009] by developing a model for group fMRI studies that can incorporate the results of previously conducted neuroimaging experiments on the same topic. This Bayesian hierarchical model performs well on simulated and real data with small sample sizes. Some increased sensitivity is found to the choice of some of the model parameters and this is a topic for further improvement. Even though the model is mainly intended for small, underpowered studies, it can be also used for larger datasets since the MCMC algorithm that is used for posterior simulation exhibits quick convergence and good mixing properties.

Finally, in Chapter 6 we have presented some initial work on estimating the extent of the file drawer problem, a major concern in CBMA but also in meta-analyses applications in other fields. Our approach uses the properties of the count distribution of reported activations and then infers the zero-count studies through a truncated model. Using data from the BrainMap database [Laird *et al.*, 2005] we estimated the prevalence of zero-count studies as roughly 1 per 10 published experiments. This estimate did not differ significantly as a function of study characteristics such as year and sample size.

## 7.2 Future work

There are several possible developments of the work presented in this dissertation. One such possibility would be to extend the log-Gaussian Cox process meta-regression model presented in Chapter 4 by allowing a homogenous random effects intensity term $\epsilon_i$ for each study. This extension can account for further variability in the observed point patterns that cannot be captured using the information from the covariates. Estimation of these parameters can be done jointly with the rest of model parameters in a single HMC step and so we expect that the increase in computation time will be moderate to low.

Regarding our group fMRI model, future work will focus on reducing sensitivity of the model to the choice of parameter $m$. One option is to consider a fourth level in the hierarchy which will contain the individual centers, as was done by Xu *et al.* [2009]. Individual centers will cluster around the study centers and will be the parent process for individual components. Therefore, changes on the total number of components will affect individuals' centers but not study centers which is the main output of the model. Another direction that we look into is the use of GPUs for the computations. Currently, all our computations are conducted in parallel using OpenMP. Even though this is efficient for levels 1 and 2 of the

model which contain few foci and subjects, we believe that there will be appreciable gains if the voxel-wise calculations of level 3 are performed with a higher degree of parallelisation.

It has been shown that image-based meta-analyses (IBMA) are superior to the kernel-based CBMA approaches [Salimi-Khorshidi *et al.*, 2009] because they contain the full information obtained from an experiment, rather than only part of it. This fact has motivated Radua *et al.* [2012] to extend the SDM method [Radua and Mataix-Cols, 2009] and model CBMA and IBMA data jointly. Currently, no such framework exists for model-based methods. We aim to address this shortcoming by proposing a meta-analysis model that borrows ideas from level 1 of the mixture model presented in Chapter 5. In particular, we will assume that for each IBMA dataset there is a latent study center process around which large statistic values are observed. The study centers will then arise from the same intensity that underlines CBMA data. Hence we can use the IBMA centers as additional foci in the CBMA analysis. Inference for this model is straightforward with a hybrid Gibbs sampler: in the first step the IBMA study centers are simulated with the spatial birth-and-death algorithm and in the second step the CBMA intensity is estimated with the HMC algorithm developed in Section 4.3.

Finally, there are some interesting open problems that relate to the file drawer problem. The zero-truncated model of Chapter 6 assumes that the zero-count studies are suppressed from the literature with probability 1. However one may argue that some of these studies change their analyses (e.g. multiple correction thresholds) until eventually some activations are obtained. We can therefore extend the zero-truncated model to allow for a small proportion $q$ of the zero-count studies to be upgraded to some non-zero count.

# Appendix A

# LGCP supplements

## A.1 Gradient expressions for the LGCP

The log-posterior, up to a normalising constant is given by:

$$\ell\left(\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \{\boldsymbol{\gamma}_k\}_{k=1}^{K^*} \mid \cdot\right) \propto \sum_{i=1}^{I}\left[-\sum_{j=1}^{V} A_{v_j}\lambda_i(v_j) + \sum_{j=1}^{V}\mathbf{1}_{v_j\in\mathbf{x}_i}\log\lambda_i(v_j)\right] + \log\text{priors}, \tag{A.1}$$

where $\boldsymbol{\beta} = [\beta_0 \equiv \mu_0, \ldots, \beta_{K^*} \equiv \mu_{K^*}, \beta_{K^*+1}, \ldots, \beta_K]^{\mathrm{T}}$, $\boldsymbol{\sigma} = [\sigma_0, \ldots, \sigma_{K^*}]^{\mathrm{T}}$, $\boldsymbol{\rho} = [\rho_0, \ldots, \rho_{K^*}]^{\mathrm{T}}$, $A_{v_j} = \mathbf{1}_{v_j\in\mathcal{B}}$ and the intensity function at each voxel $v_j$ is defined as:

$$\lambda_i(v_j) = \exp\left(\sum_{k=0}^{K}\beta_k z_{ik}\right)\exp\left(\sum_{k=0}^{K^*}\sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik}\right) \tag{A.2}$$

We now calculate the derivatives with respect to the parameters of interest.

### Partial derivatives with respect to $\beta_l$

We have that:

$$\begin{aligned}\frac{\partial\log\lambda_i(v_j)}{\partial\beta_l} &= \frac{\partial}{\partial\beta_l}\sum_{k=0}^{K}\beta_k z_{ik} + \frac{\partial}{\partial\beta_l}\sum_{k=0}^{K^*}\sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik}\\ &= z_{il}.\end{aligned} \tag{A.3}$$

As a result:

$$\frac{\partial\ell\left(\beta_l \mid \cdot\right)}{\partial\beta_l} = -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\frac{\partial}{\partial\beta_l}\lambda_i(v_j) - \mathbf{1}_{v_j\in\mathbf{x}_i}\frac{\partial}{\partial\beta_l}\log\lambda_i(v_j)\right] + \frac{\partial}{\partial\beta_l}\log\pi(\beta_l)$$

$$
\begin{aligned}
&= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}z_{il}\right] - \frac{\partial}{\partial\beta_l}\frac{\beta_l^2}{2\tau^2} \\
&= -\sum_{j=1}^{V}\sum_{i=1}^{I}\left[A_{v_j}\lambda_i(v_j)z_{il} + n_i z_{il}\right] - \frac{\beta_l}{\tau^2}, \qquad \text{(A.4)}
\end{aligned}
$$

where $n_i$ is the total number of foci in study $i$.

### Partial derivatives with respect to $\sigma_l$

We have that:

$$
\begin{aligned}
\frac{\partial \log\lambda_i(v_j)}{\partial\sigma_l} &= \frac{\partial}{\partial\sigma_l}\sum_{k=0}^{K}\beta_k z_{ik} + \frac{\partial}{\partial\sigma_l}\sum_{k=0}^{K^*}\sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik} \\
&= \left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik}. \qquad \text{(A.5)}
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
\frac{\partial\ell\left(\sigma_l \mid \cdot\right)}{\partial\sigma_l} &= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\frac{\partial}{\partial\sigma_l}\lambda_i(v_j) - \mathbf{1}_{v_j\in\mathbf{x}_i}\frac{\partial}{\partial\sigma_l}\log\lambda_i(v_j)\right] + \frac{\partial}{\partial\sigma_l}\log\pi(\sigma_l) \\
&= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{vj}\lambda_i(v_j)\left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}\left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j z_{il}\right] - \frac{\partial}{\partial\sigma_l}\frac{\sigma_l^2}{2\tau^2} \\
&= -\sum_{j=1}^{V}\sum_{i=1}^{I}\left[\left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j\left(A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}z_{il}\right)\right] - \frac{\sigma_l}{\tau^2} \\
&= -\sum_{j=1}^{V}\left[\left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j\sum_{i=1}^{I}\left[A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}z_{il}\right]\right] - \frac{\sigma_l}{\tau^2}.
\end{aligned}
$$

### Partial derivatives with respect to $\rho_l$

Again:

$$
\begin{aligned}
\frac{\partial \log\lambda_i(v_j)}{\partial\rho_l} &= \frac{\partial}{\partial\rho_l}\sum_{k=0}^{K^*}\beta_k z_{ik} + \frac{\partial}{\partial\rho_l}\sum_{k=0}^{K}\sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik} \\
&= \sigma_l\frac{\partial}{\partial\rho_l}\left(\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l\right)_j z_{il}. \qquad \text{(A.6)}
\end{aligned}
$$

For ease of exposition we will complete the derivation for the one-dimensional case; however, similar arguments hold when $\mathcal{B}\subset\mathbb{R}^3$. Matrices $\mathbf{R}_l$ are circulant and so, the matrix-vector product $\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l$ can be found using the discrete Fourier transform

as:

$$\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l = \mathbf{F}\boldsymbol{\Phi}_l^{1/2}\mathbf{F}^{\mathrm{H}}\boldsymbol{\gamma}_l, \tag{A.7}$$

where $\boldsymbol{\Phi}_l$ are the diagonal matrices containing the eigenvalues of $\mathbf{R}_l$ and $\mathbf{F}$ is the matrix of eigenvectors. In Equation (A.7), the only term depending on $\rho_l$ is $\boldsymbol{\Phi}_l$ and, hence:

$$\frac{\partial}{\partial\rho_l}\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l = \mathbf{F}\frac{\partial}{\partial\rho_l}\boldsymbol{\Phi}_l^{1/2}\mathbf{F}^{\mathrm{H}}\boldsymbol{\gamma}_l \tag{A.8}$$

We know that $\boldsymbol{\Phi}_l = \mathrm{diag}\left\{\phi_{l_0},\ldots,\phi_{l_{V-1}}\right\}$, where for $k = 0,\ldots,V-1$ we have that:

$$\phi_{l_k} = \sum_{j=0}^{V-1}\exp\left(-\rho_l||v_0,v_j||^{\delta_l}\right)\exp\left(-\frac{2\pi\iota kj}{V}\right), \tag{A.9}$$

where $\iota$ is the imaginary unit. Now it is straightforward to see that for $k = 0,\ldots,V-1$:

$$\begin{aligned}
\frac{\partial}{\partial\rho_l}\phi_{l_k}^{1/2} &= \frac{\partial}{\partial\rho_l}\sqrt{\sum_{j=0}^{V-1}\exp\left(-\rho_l||v_0,v_j||^{\delta_l}\right)\exp\left(-\frac{2\pi\iota kj}{V}\right)} \\
&= \frac{\frac{\partial}{\partial\rho_l}\sum_{j=0}^{V-1}\exp\left(-\rho_l||v_0,v_j||^{\delta_l}\right)\exp\left(-\frac{2\pi\iota kj}{V}\right)}{2\sqrt{\sum_{j=0}^{V-1}\exp\left(-\rho_l||v_0,v_j||^{\delta_l}\right)\exp\left(-\frac{2\pi\iota kj}{V}\right)}} \\
&= \frac{-\sum_{j=1}^{V-1}d\left(v_0,v_j\right)^{\delta_l}\exp\left(-\rho_l||v_0,v_j||^{\delta_l}\right)\exp\left(-\frac{2\pi\iota kj}{V}\right)}{2\phi_{l_k}^{1/2}} \\
&= -\frac{1}{2}\frac{\psi_{l_k}}{\phi_{l_k}^{1/2}},
\end{aligned} \tag{A.10}$$

where $\psi_{l_k}$ can be viewed as the $k$-th eigenvalue of the of a circulant matrix $\mathbf{S}_l$ with base $\mathbf{s}_l = \left[||v_0,v_0||^{\delta_l}\exp\left(-\rho_l||v_0,v_0||^{\delta_l}\right),\ldots,||v_0,v_{V-1}||^{\delta_l}\exp\left(-\rho_l||v_0,v_{V-1}||^{\delta_l}\right)\right]$ and $\mathbf{S}_l = \mathbf{F}\boldsymbol{\Psi}_l\mathbf{F}^{\mathrm{H}}$, $\boldsymbol{\Psi}_l = \mathrm{diag}\left\{\psi_{l_0},\ldots,\psi_{l_{V-1}}\right\}$. Overall we see that:

$$\begin{aligned}
\frac{\partial}{\partial\rho_l}\mathbf{R}_l^{1/2}\boldsymbol{\gamma}_l &= \mathbf{F}\frac{\partial}{\partial\rho_l}\boldsymbol{\Phi}_l^{1/2}\mathbf{F}^{\mathrm{H}}\boldsymbol{\gamma}_l \\
&= -\frac{1}{2}\mathbf{F}\left[\boldsymbol{\Psi}_l\oslash\boldsymbol{\Phi}^{1/2}\right]\mathbf{F}^{\mathrm{H}}\boldsymbol{\gamma}_l \\
&= -\frac{1}{2}\mathbf{Q}_l\boldsymbol{\gamma}_l,
\end{aligned} \tag{A.11}$$

where $\oslash$ stands for element wise division. Combining Equations (A.6) and (A.11), we find that:

$$\frac{\partial \log \lambda_i(v_j)}{\partial \rho_l} = -\frac{1}{2}\sigma_l \left(\mathbf{Q}_l \boldsymbol{\gamma}_l\right)_j z_{il}. \tag{A.12}$$

So:

$$\begin{aligned}
\frac{\partial \ell\left(\rho_l \mid \cdot\right)}{\partial \rho_l} &= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\frac{\partial}{\partial \rho_l}\lambda_i(v_j) - \mathbf{1}_{v_j \in \mathbf{x}_i}\frac{\partial}{\partial \rho_l}\log \lambda_i(v_j)\right] + \frac{\partial}{\partial \rho_l}\log \pi(\rho_l) \\
&= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{vj}\lambda_i(v_j)\left(-\frac{1}{2}\right)\sigma_l\left(\mathbf{Q}_l\boldsymbol{\gamma}_l\right)_j z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}\left(-\frac{1}{2}\right)\sigma_l\left(\mathbf{Q}_l\boldsymbol{\gamma}_l\right)_j z_{il}\right] \\
&\quad -\frac{\partial}{\partial \rho_l}1 \\
&= \frac{\sigma_l}{2}\sum_{j=1}^{V}\sum_{i=1}^{I}\left[\left(\mathbf{Q}_l\boldsymbol{\gamma}_l\right)_j\left(A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}z_{il}\right)\right] \\
&= \frac{\sigma_l}{2}\sum_{j=1}^{V}\left[\left(\mathbf{Q}_l\boldsymbol{\gamma}_l\right)_j\sum_{i=1}^{I}\left[A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}z_{il}\right]\right].
\end{aligned}$$

### Partial derivatives with respect to $\boldsymbol{\gamma}_l$

Finally:

$$\begin{aligned}
\frac{\partial \log \lambda_i(v_j)}{\partial \boldsymbol{\gamma}_l} &= \frac{\partial}{\partial \boldsymbol{\gamma}_l}\sum_{k=0}^{K^*}\beta_k z_{ik} + \frac{\partial}{\partial \boldsymbol{\gamma}_l}\sum_{k=0}^{K}\sigma_k\left(\mathbf{R}_k^{1/2}\boldsymbol{\gamma}_k\right)_j z_{ik} \\
&= \sigma_l \mathbf{r}_{l_j} z_{il},
\end{aligned} \tag{A.13}$$

where $\mathbf{r}_{l_j}$ is the $j$-th row of the matrix $\mathbf{R}_l^{1/2}$. Now we can see that:

$$\begin{aligned}
\frac{\partial \ell\left(\boldsymbol{\gamma}_l \mid \cdot\right)}{\partial \boldsymbol{\gamma}_l} &= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\frac{\partial}{\partial \boldsymbol{\gamma}_l}\lambda_i(v_j) - \mathbf{1}_{v_j\in\mathbf{x}_i}\frac{\partial}{\partial \boldsymbol{\gamma}_l}\log\lambda_i(v_j)\right] + \frac{\partial}{\partial \boldsymbol{\gamma}_l}\log\pi(\boldsymbol{\gamma}_l) \\
&= -\sum_{i=1}^{I}\sum_{j=1}^{V}\left[A_{v_j}\lambda_i(v_j)\sigma_l\mathbf{r}_{l_j}z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}\sigma_l\mathbf{r}_{l_j}z_{il}\right] - \frac{\partial}{\partial \boldsymbol{\gamma}_l}\frac{\boldsymbol{\gamma}_l^{\mathrm{T}}\boldsymbol{\gamma}_l}{2} \\
&= -\sigma_l\sum_{j=1}^{V}\left[\mathbf{r}_{l_j}\sum_{i=1}^{I}\left[A_{v_j}\lambda_i(v_j)z_{il} - \mathbf{1}_{v_j\in\mathbf{x}_i}z_{il}\right]\right] - \boldsymbol{\gamma}_l \\
&= -\sigma_l\sum_{j=1}^{V}\left[\mathbf{r}_{l_j}\mathbf{c}_{l_j}\right] - \boldsymbol{\gamma}_l \\
&= -\sigma_l{\mathbf{R}_l^{1/2}}^{\mathrm{T}}\mathbf{c}_l - \boldsymbol{\gamma}_l
\end{aligned}$$

$$= \quad -\sigma_l \mathbf{R}_l^{1/2} \mathbf{c}_l - \boldsymbol{\gamma}_l, \tag{A.14}$$

since $\mathbf{R}$ is a nested block circulant matrix, where $\mathbf{c}_l$ are $V$-vectors with elements $\mathbf{c}_{l_j} = \sum_{i=1}^{I} \left[ A_{v_j} \lambda_i(v_j) z_{il} - \mathbf{1}_{v_j \in \mathbf{x_i}} z_{il} \right]$.

## A.2 LGCP simulation setup I traceplots

In this section we provide traceplots for the scalar parameters of the LGCP model, as fit to the simulated data of Section 4.4.1. Trace plots for the parameters $\sigma_1$, $\sigma_2$, $\rho_1$ and $\rho_2$ can be found in Figure A.1, whereas trace plots for $\mu_1$, $\mu_2$, $\beta_3$ and $\beta_4$ can be found in Figure A.2. The red lines indicate the true parameter values.

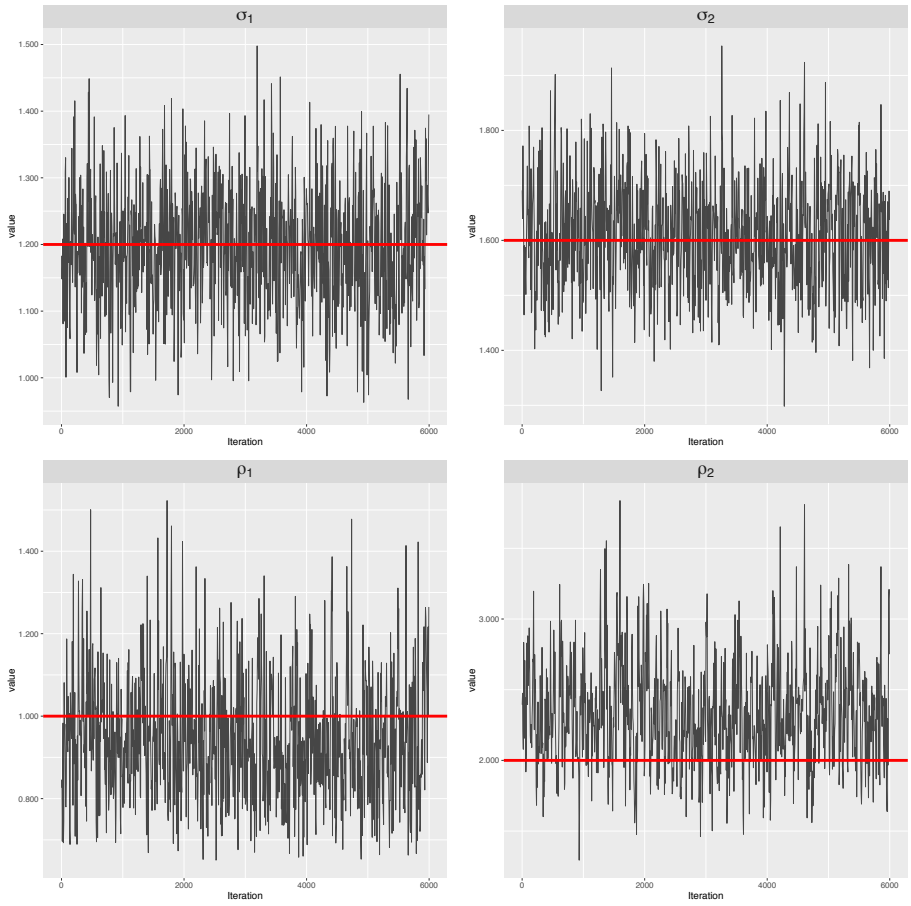



Figure A.1: Posterior traceplots for the scalar parameters of the LGCP model used to fit the data of Section 4.4.1. Top row: standard deviations. Bottom row: correlation decay parameters (×100). The true values are indicated by the solid red lines.

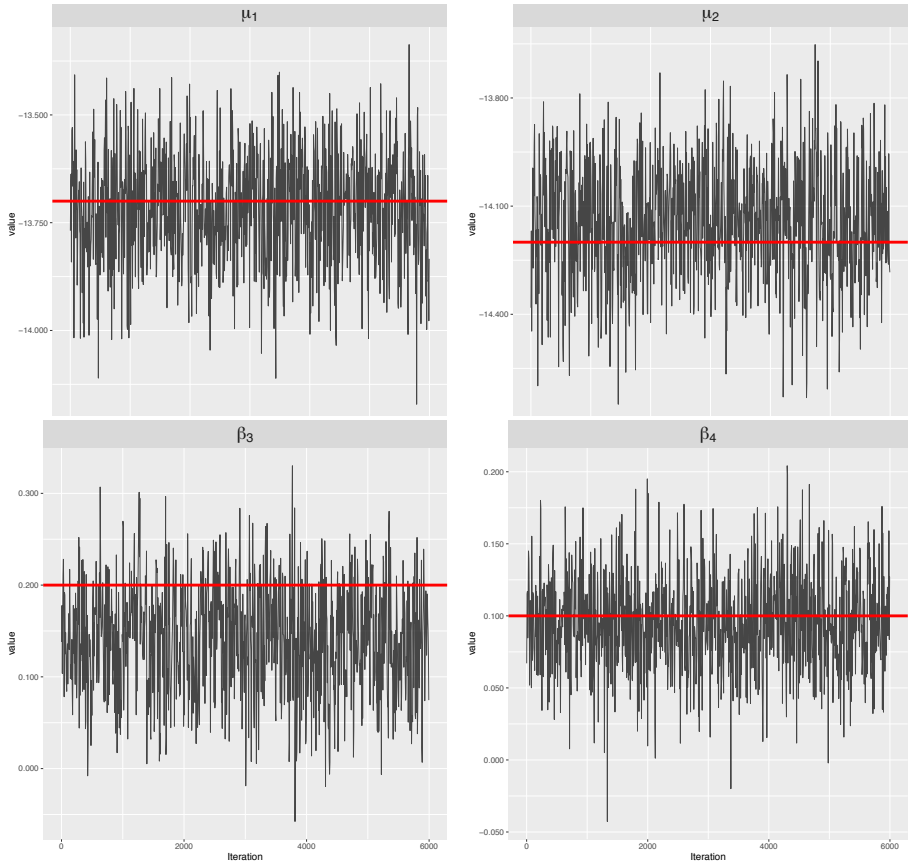



Figure A.2: Posterior traceplots for the scalar parameters of the LGCP model used to fit the data of Section 4.4.1. Top row: overall latent field means. Bottom row: regression coefficients for covariates $z_3$ and $z_4$. The true values are indicated by the solid red lines.

## A.3 Real data analysis diagnostics

This section contains convergence diagnostics for the real data analysis of Section 4.5. All plots are obtained from 3 different runs of HMC, from which we discard the first 5,000 draws as a burn-in. Parameters $\sigma_k$, $\rho_k$ and $\mu_k$, $k = 1, 2$, are shown in Figures A.3, A.4 and A.5 respectively. Figure A.6 shows draws from the posterior expected number of points per type. In Figure A.7, we plot posterior intensities for voxel $v_1$ where the highest median posterior intensity for emotions is observed. Finally, in Figure A.8 we do the same for voxel $v_2$ which corresponds to the highest median posterior intensity for executive control studies .

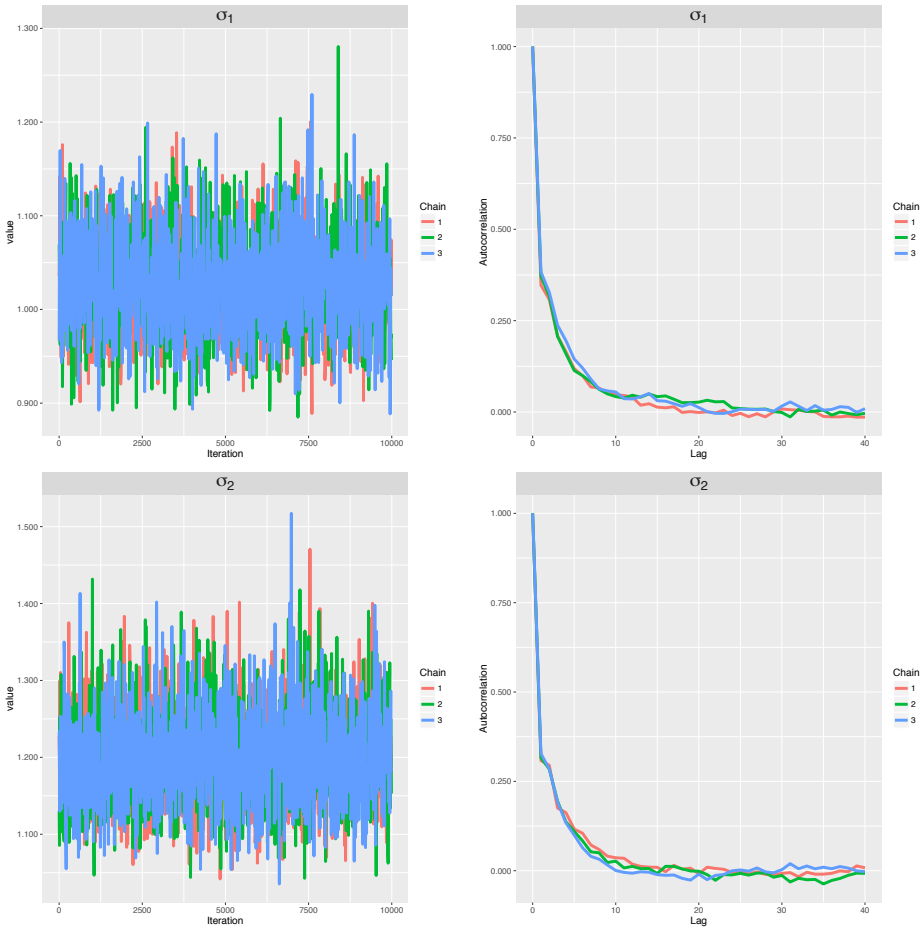



Figure A.3: Posterior traceplots (left) along with autocorrelation plots (right) for the marginal standard deviations $\sigma_k$, as obtained from application of the LGCP meta-analysis model to the real data of Section 4.5. Top row corresponds to emotion while bottom row to executive control.

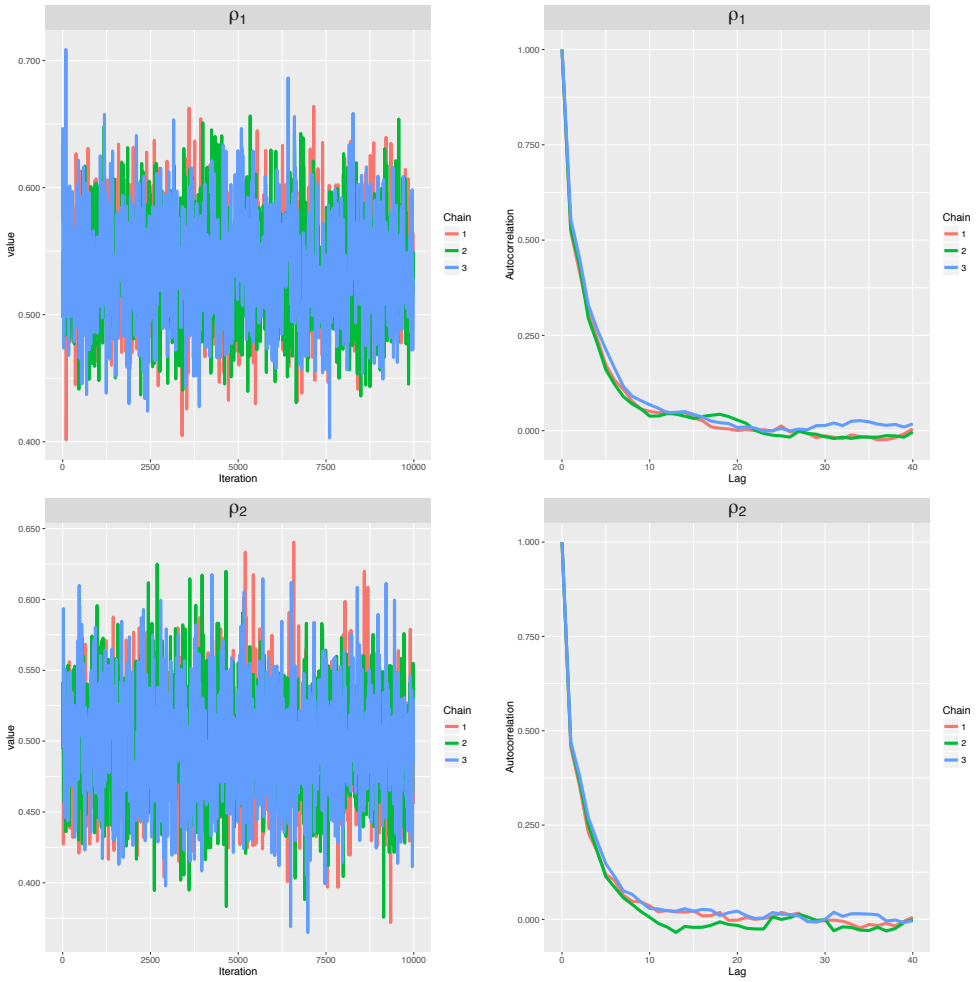



Figure A.4: Posterior traceplots (left) along with autocorrelation plots (right) for the correlation decay parameters $\rho_k$, as obtained from application of the LGCP meta-analysis model to the real data of Section 4.5. Top row corresponds to emotion while bottom row to executive control. The actual values are multiplied by 100.

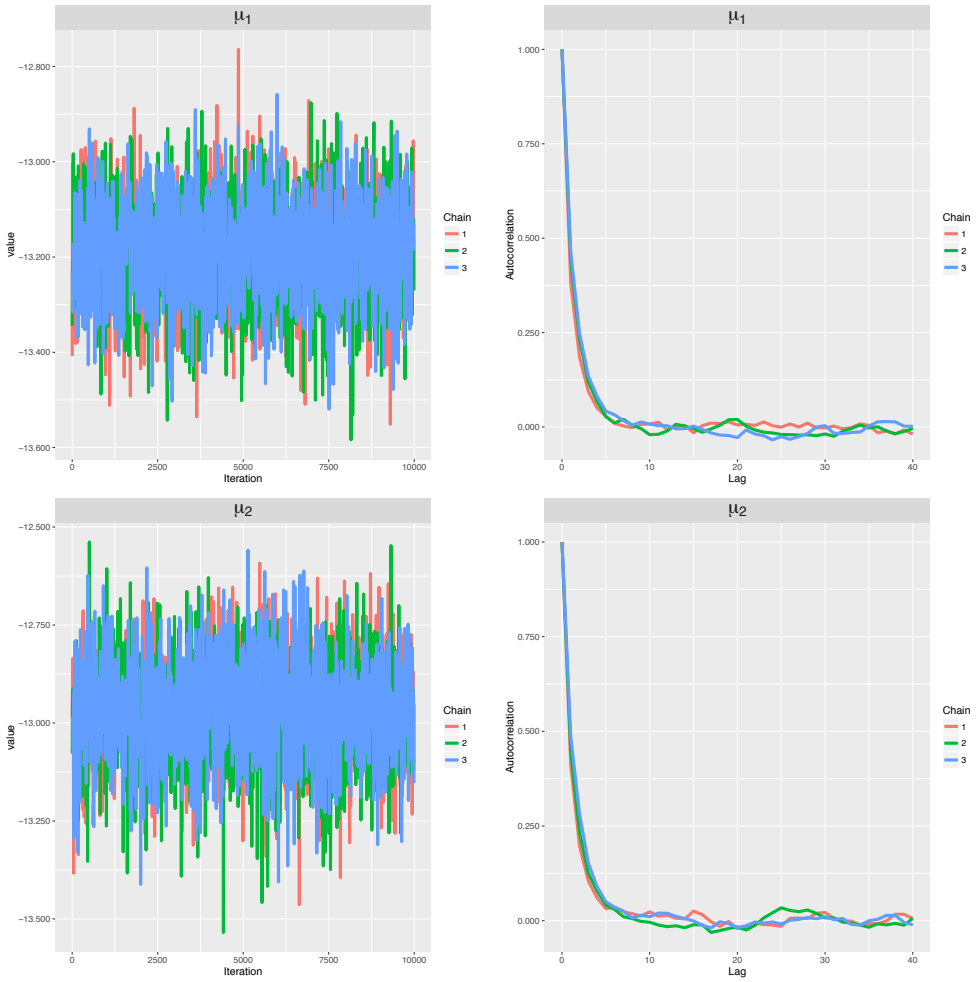



Figure A.5: Posterior traceplots (left) along with autocorrelation plots (right) for the overall latent field means $\mu_k$, as obtained from application of the LGCP meta-analysis model to the real data of Section 4.5. Top row corresponds to emotion while bottom row to executive control.

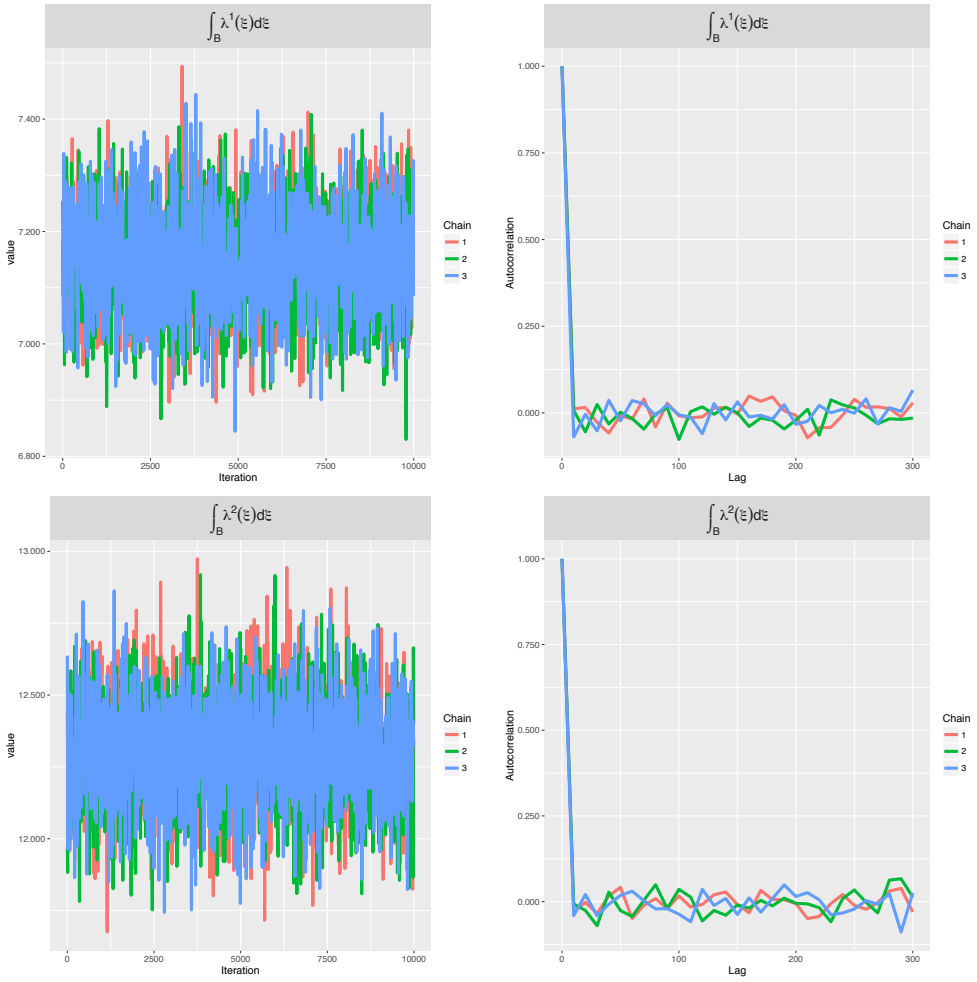



Figure A.6: Posterior traceplots (left) along with autocorrelation plots (right) for the integrated intensities over the brain, as obtained from application of the LGCP meta-analysis model to the real data of Section 4.5. Top row corresponds to emotion while bottom row to executive control. Results are obtained after applying a thinning factor of 10 to the original draws.

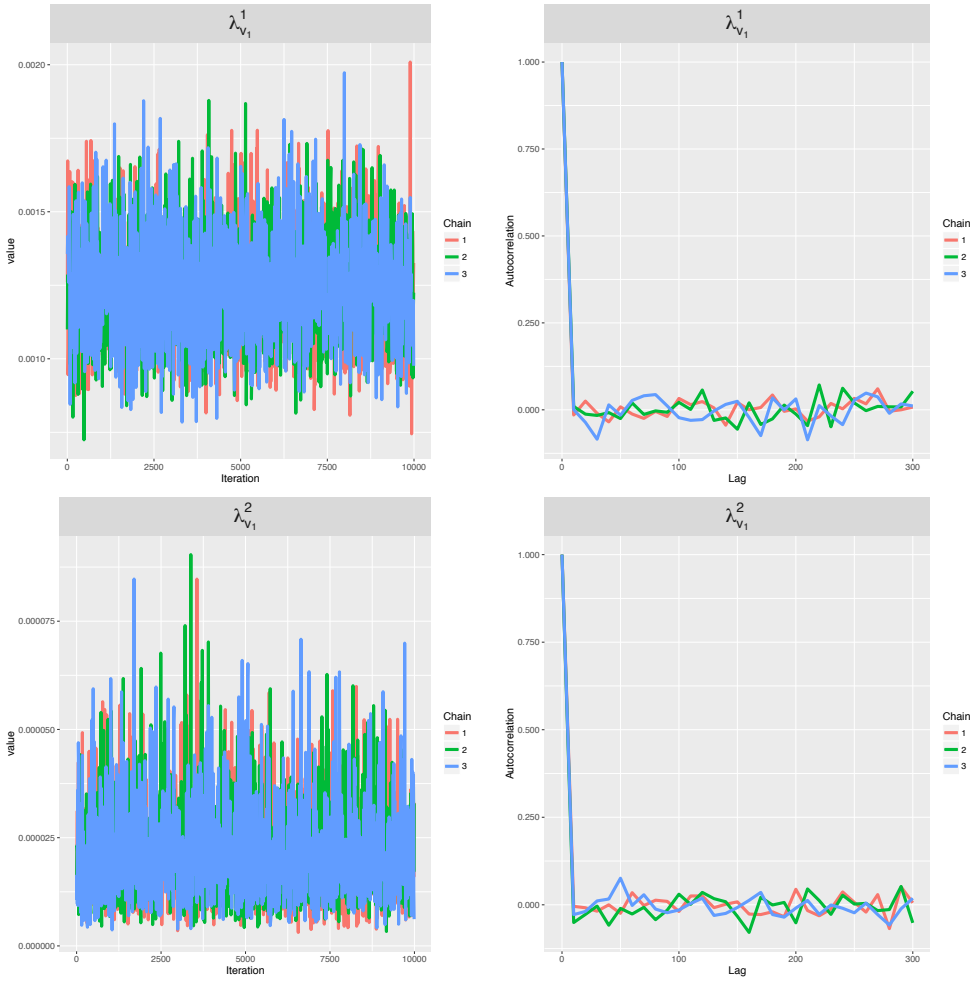



Figure A.7: Posterior traceplots (left) along with autocorrelation plots (right) for intensities at voxel $v_1$ where the highest median posterior intensity was observed for emotions, as obtained from application of the LGCP meta-analysis model to the real data of Section 4.5. Top row corresponds to emotion while bottom row to executive control. Results are obtained after applying a thinning factor of 10 to the original draws.

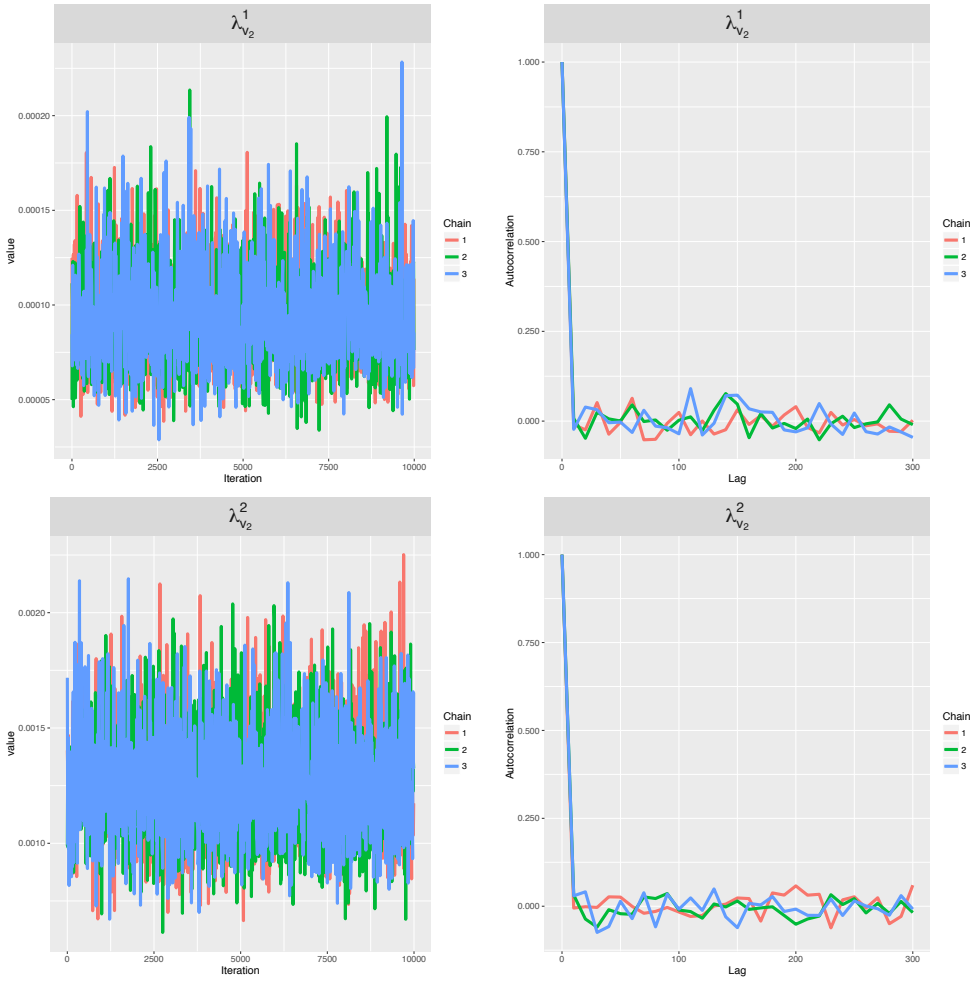



Figure A.8: Posterior traceplots (left) along with autocorrelation plots (right) for intensities at voxel $v_2$ where the highest median posterior intensity was observed for executive control, as obtained from application of the LGCP meta-analysis model to the real data of Section 4.5. Top row corresponds to emotion while bottom row to executive control. Results are obtained after applying a thinning factor of 10 to the original draws.

## A.4 Full brain analysis

In this section we present full brain results for the application of Section 4.5. Table A.1 show posterior summaries for emotions whereas Table A.2 shows posterior summaries for executive control. ROIs are obtained from the Harvard-Oxford atlas [Desikan *et al.*, 2006]. The ROI volumes presented are in voxels. All quantities are based on 1,000 HMC samples, obtained after thinning the original chain every 10 iterations.

Table A.1: Meta-analysis results for emotions. Table presents posterior summaries for the probability of at least one activation in a given ROI (rounded to 2 decimal points), as well as the expected number points (rounded to 3 decimal points). The empirical quantities are also presented. Note the abbreviations: (L) for left, (R) for right, (A) for anterior, (P) for posterior, (S) for superior and (I) for inferior. All results are based on 1,000 HMC draws which are obtained after thinning the initial run every 10 iterations.

| | | $\% \, \mathbb{P}\left(\mathbf{N_X}\left(\mathbf{B}\right) \geq \mathbf{1}\right)$ | | | | $\int_{\mathbf{B}} \lambda\left(\xi\right) \mathbf{d}\xi$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| **ROI** | **VOL.** | $\mathbf{p_{0.5}}$ | $\mathbf{p_{0.025}}$ | $\mathbf{p_{0.975}}$ | **DATA** | $\mathbf{p_{0.5}}$ | $\mathbf{p_{0.025}}$ | $\mathbf{p_{0.975}}$ | **DATA** |
| Lateral ventrical (L) | 1289 | 2.51 | 1.96 | 3.17 | 2.34 | 0.025 | 0.020 | 0.032 | 0.028 |
| Lateral ventrical (R) | 1019 | 1.99 | 1.54 | 2.50 | 1.64 | 0.020 | 0.016 | 0.025 | 0.020 |
| Thalamus (L) | 1591 | 10.48 | 8.85 | 12.40 | 10.29 | 0.111 | 0.093 | 0.132 | 0.120 |
| Thalamus (R) | 1398 | 7.97 | 6.53 | 9.61 | 8.65 | 0.083 | 0.068 | 0.101 | 0.092 |
| Caudate (L) | 572 | 2.04 | 1.53 | 2.74 | 2.11 | 0.021 | 0.015 | 0.028 | 0.022 |
| Caudate (R) | 515 | 1.70 | 1.23 | 2.29 | 1.99 | 0.017 | 0.012 | 0.023 | 0.022 |
| Putamen (L) | 923 | 4.49 | 3.48 | 5.64 | 4.91 | 0.046 | 0.035 | 0.058 | 0.051 |
| Putamen (R) | 800 | 3.58 | 2.85 | 4.52 | 3.86 | 0.036 | 0.029 | 0.046 | 0.042 |
| Pallidum (L) | 312 | 1.42 | 1.01 | 1.95 | 0.70 | 0.014 | 0.010 | 0.020 | 0.007 |
| Pallidum (R) | 266 | 1.61 | 1.19 | 2.21 | 1.52 | 0.016 | 0.012 | 0.022 | 0.016 |
| Brain stem | 7413 | 11.56 | 9.96 | 13.34 | 9.24 | 0.123 | 0.105 | 0.143 | 0.120 |
| Hippocampus (L) | 921 | 11.11 | 9.52 | 12.77 | 11.35 | 0.118 | 0.100 | 0.137 | 0.123 |
| Hippocampus (R) | 772 | 8.95 | 7.56 | 10.45 | 7.72 | 0.094 | 0.079 | 0.110 | 0.087 |
| Amygdala (L) | 390 | 13.87 | 12.02 | 15.78 | 15.09 | 0.149 | 0.128 | 0.172 | 0.159 |
| Amygdala (R) | 399 | 12.46 | 10.69 | 14.42 | 13.22 | 0.133 | 0.113 | 0.156 | 0.138 |
| Accumbens (L) | 111 | 0.60 | 0.39 | 0.90 | 1.05 | 0.006 | 0.004 | 0.009 | 0.011 |
| Accumbens (R) | 86 | 0.41 | 0.27 | 0.62 | 0.47 | 0.004 | 0.003 | 0.006 | 0.005 |
| Frontal pole | 15188 | 29.16 | 26.87 | 31.76 | 23.04 | 0.345 | 0.313 | 0.382 | 0.331 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Insular cortex | 3613 | 19.50 | 17.63 | 21.73 | 15.09 | 0.217 | 0.194 | 0.245 | 0.226 |
| Superior frontal gyrus | 7295 | 18.13 | 16.19 | 20.26 | 14.97 | 0.200 | 0.177 | 0.226 | 0.204 |
| Middle frontal gyrus | 6722 | 12.46 | 10.85 | 14.34 | 9.47 | 0.133 | 0.115 | 0.155 | 0.113 |
| Inferior frontal gyrus, pars triangularis | 1615 | 10.05 | 8.59 | 11.62 | 9.94 | 0.106 | 0.090 | 0.124 | 0.109 |
| Inferior frontal gyrus, pars opercularis | 1877 | 14.35 | 12.55 | 16.23 | 12.98 | 0.155 | 0.134 | 0.177 | 0.172 |
| Precentral gyrus | 11055 | 25.76 | 23.44 | 28.29 | 20.70 | 0.298 | 0.267 | 0.333 | 0.322 |
| Temporal pole | 3341 | 17.78 | 15.73 | 20.02 | 14.97 | 0.196 | 0.171 | 0.223 | 0.192 |
| Superior temporal gyrus (A) | 622 | 2.20 | 1.69 | 2.87 | 2.11 | 0.022 | 0.017 | 0.029 | 0.025 |
| Superior temporal gyrus (P) | 2093 | 8.83 | 7.41 | 10.17 | 7.25 | 0.092 | 0.077 | 0.107 | 0.090 |
| Middle temporal gyrus (A) | 762 | 2.73 | 2.05 | 3.63 | 2.69 | 0.028 | 0.021 | 0.037 | 0.028 |
| Middle temporal gyrus (P) | 2794 | 8.37 | 6.98 | 9.83 | 7.60 | 0.087 | 0.072 | 0.103 | 0.094 |
| Middle temporooccipital temporal gyrus | 2470 | 10.48 | 8.93 | 12.17 | 9.47 | 0.111 | 0.094 | 0.130 | 0.119 |
| Inferior temporal gyrus (A) | 789 | 1.18 | 0.80 | 1.69 | 1.05 | 0.012 | 0.008 | 0.017 | 0.012 |
| Inferior temporal gyrus (P) | 2687 | 3.08 | 2.33 | 3.99 | 1.99 | 0.031 | 0.024 | 0.041 | 0.020 |
| Inferior temporoocipital temporal gyrus | 1883 | 5.06 | 4.11 | 6.20 | 3.86 | 0.052 | 0.042 | 0.064 | 0.042 |
| Postcentral gyrus | 7553 | 12.30 | 10.67 | 14.09 | 8.19 | 0.131 | 0.113 | 0.152 | 0.112 |
| Superior parietal lobule | 3626 | 4.91 | 3.86 | 6.08 | 4.21 | 0.050 | 0.039 | 0.063 | 0.047 |
| Supramarginal gyrus, (A) | 1951 | 4.30 | 3.35 | 5.49 | 4.21 | 0.044 | 0.034 | 0.056 | 0.051 |
| Supramarginal gyrus, (P) | 3017 | 6.10 | 5.04 | 7.31 | 5.50 | 0.063 | 0.052 | 0.076 | 0.063 |
| Angular gyrus | 2944 | 5.98 | 4.86 | 7.25 | 5.15 | 0.062 | 0.050 | 0.075 | 0.058 |
| Lateral occipital cortex (S) | 9535 | 17.99 | 15.92 | 20.27 | 13.45 | 0.198 | 0.173 | 0.226 | 0.198 |
| Lateral occipital cortex (I) | 5004 | 30.57 | 28.02 | 33.29 | 22.46 | 0.365 | 0.329 | 0.405 | 0.382 |
| Intracalcarine cortex | 2211 | 4.89 | 3.92 | 6.09 | 4.44 | 0.050 | 0.040 | 0.063 | 0.048 |
| Frontal medial cortex | 1473 | 6.35 | 5.18 | 7.72 | 5.50 | 0.066 | 0.053 | 0.080 | 0.061 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Juxtapositional lobule cortex | 2193 | 7.97 | 6.56 | 9.59 | 8.30 | 0.083 | 0.068 | 0.101 | 0.091 |
| Subcallosal cortex | 2150 | 7.21 | 5.81 | 8.81 | 5.96 | 0.075 | 0.060 | 0.092 | 0.078 |
| Paracingulate gyrus | 4095 | 18.71 | 16.74 | 20.88 | 15.67 | 0.207 | 0.183 | 0.234 | 0.202 |
| Cingulate gyrus (A) | 4144 | 20.39 | 18.30 | 22.58 | 17.78 | 0.228 | 0.202 | 0.256 | 0.237 |
| Cingulate gyrus (P) | 4668 | 10.22 | 8.79 | 11.97 | 8.54 | 0.108 | 0.092 | 0.128 | 0.101 |
| Precuneous cortex | 7439 | 14.43 | 12.53 | 16.58 | 12.05 | 0.156 | 0.134 | 0.181 | 0.158 |
| Cuneal cortex | 1605 | 2.06 | 1.52 | 2.80 | 1.40 | 0.021 | 0.015 | 0.028 | 0.015 |
| Frontal orbital cortex | 4058 | 25.06 | 22.81 | 27.46 | 21.40 | 0.289 | 0.259 | 0.321 | 0.298 |
| Parahippocampal gyrus (A) | 3225 | 28.56 | 26.20 | 30.82 | 23.63 | 0.336 | 0.304 | 0.368 | 0.332 |
| Parahippocampal gyrus (P) | 2014 | 7.72 | 6.41 | 9.17 | 7.02 | 0.080 | 0.066 | 0.096 | 0.089 |
| Lingual gyrus | 5387 | 14.04 | 12.35 | 16.01 | 11.35 | 0.151 | 0.132 | 0.174 | 0.156 |
| Temporal fusiform cortex (A) | 830 | 1.40 | 1.00 | 1.87 | 1.05 | 0.014 | 0.010 | 0.019 | 0.013 |
| Temporal fusiform cortex (P) | 2728 | 5.45 | 4.53 | 6.57 | 4.91 | 0.056 | 0.046 | 0.068 | 0.055 |
| Temporal occipital fusiform cortex | 2451 | 17.51 | 15.48 | 19.63 | 15.56 | 0.193 | 0.168 | 0.219 | 0.212 |
| Occipital fusiform gyrus | 3576 | 14.36 | 12.58 | 16.30 | 11.81 | 0.155 | 0.134 | 0.178 | 0.146 |
| Frontal operculum cortex | 1059 | 6.91 | 5.84 | 8.03 | 5.85 | 0.072 | 0.060 | 0.084 | 0.065 |
| Central operculum cortex | 2568 | 7.54 | 6.46 | 8.83 | 6.32 | 0.078 | 0.067 | 0.092 | 0.070 |
| Parietal operculum cortex | 1683 | 3.35 | 2.55 | 4.27 | 3.39 | 0.034 | 0.026 | 0.044 | 0.039 |
| Planum polare | 1210 | 4.27 | 3.54 | 5.10 | 3.63 | 0.044 | 0.036 | 0.052 | 0.041 |
| Heschl's gyrus | 786 | 2.90 | 2.21 | 3.72 | 3.51 | 0.029 | 0.022 | 0.038 | 0.039 |
| Planum temporale | 1434 | 4.53 | 3.68 | 5.39 | 4.44 | 0.046 | 0.037 | 0.055 | 0.049 |
| Supracalcarine cortex | 424 | 0.78 | 0.55 | 1.08 | 0.94 | 0.008 | 0.005 | 0.011 | 0.009 |
| Occipital pole | 4081 | 15.36 | 13.27 | 17.56 | 12.40 | 0.167 | 0.142 | 0.193 | 0.175 |

Table A.2: Meta-analysis results for executive control. Table presents posterior summaries for the probability of at least one activation in a given ROI (rounded to 2 decimal points), as well as the expected number points (rounded to 3 decimal points). The empirical quantities are also presented. Note the abbreviations: (L) for left, (R) for right, (A) for anterior, (P) for posterior, (S) for superior and (I) for inferior. All results are based on 1,000 HMC draws which are obtained after thinning the initial run every 10 iterations.

| | | $\% \, \mathbb{P}\left(\mathbf{N_X}\left(\mathbf{B}\right) \geq \mathbf{1}\right)$ | | | | $\int_{\mathbf{B}} \lambda\left(\xi\right) \mathbf{d}\xi$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| **ROI** | **VOL.** | $\mathbf{p_{0.5}}$ | $\mathbf{p_{0.025}}$ | $\mathbf{p_{0.975}}$ | **DATA** | $\mathbf{p_{0.5}}$ | $\mathbf{p_{0.025}}$ | $\mathbf{p_{0.975}}$ | **DATA** |
| Lateral ventrical (L) | 1289 | 3.48 | 2.57 | 4.81 | 2.96 | 0.035 | 0.026 | 0.049 | 0.030 |
| Lateral ventrical (R) | 1019 | 2.65 | 1.90 | 3.54 | 2.37 | 0.027 | 0.019 | 0.036 | 0.024 |
| Thalamus (L) | 1591 | 14.65 | 11.84 | 18.26 | 13.02 | 0.158 | 0.126 | 0.202 | 0.154 |
| Thalamus (R) | 1398 | 11.56 | 8.89 | 14.41 | 10.95 | 0.123 | 0.093 | 0.156 | 0.127 |
| Caudate (L) | 572 | 3.39 | 2.36 | 4.73 | 3.25 | 0.034 | 0.024 | 0.048 | 0.033 |
| Caudate (R) | 515 | 2.60 | 1.73 | 3.65 | 2.37 | 0.026 | 0.017 | 0.037 | 0.024 |
| Putamen (R) | 800 | 4.73 | 3.36 | 6.44 | 5.62 | 0.048 | 0.034 | 0.067 | 0.059 |
| Putamen (L) | 923 | 5.69 | 4.21 | 7.54 | 5.33 | 0.059 | 0.043 | 0.078 | 0.053 |
| Pallidum (L) | 312 | 3.37 | 2.29 | 4.85 | 3.25 | 0.034 | 0.023 | 0.050 | 0.033 |
| Pallidum (R) | 266 | 2.48 | 1.56 | 3.71 | 2.66 | 0.025 | 0.016 | 0.038 | 0.027 |
| Brain stem | 7413 | 8.18 | 5.90 | 10.73 | 7.10 | 0.085 | 0.061 | 0.113 | 0.080 |
| Hippocampus (L) | 921 | 2.36 | 1.54 | 3.52 | 2.07 | 0.024 | 0.016 | 0.036 | 0.024 |
| Hippocampus (R) | 772 | 1.26 | 0.75 | 2.05 | 1.48 | 0.013 | 0.008 | 0.021 | 0.018 |
| Amygdala (L) | 390 | 0.84 | 0.43 | 1.48 | 0.59 | 0.008 | 0.004 | 0.015 | 0.006 |
| Amygdala (R) | 399 | 0.40 | 0.18 | 0.81 | 0.30 | 0.004 | 0.002 | 0.008 | 0.003 |
| Accumbens (L) | 111 | 0.52 | 0.27 | 1.00 | 1.48 | 0.005 | 0.003 | 0.010 | 0.015 |
| Accumbens (R) | 86 | 0.40 | 0.20 | 0.79 | 0.00 | 0.004 | 0.002 | 0.008 | 0.000 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Frontal pole | 15188 | 55.22 | 51.03 | 59.15 | 45.27 | 0.803 | 0.714 | 0.895 | 0.796 |
| Insular cortex | 3613 | 32.71 | 28.90 | 36.61 | 30.47 | 0.396 | 0.341 | 0.456 | 0.450 |
| Superior frontal gyrus | 7295 | 30.85 | 27.48 | 34.54 | 25.15 | 0.369 | 0.321 | 0.424 | 0.317 |
| Middle frontal gyrus | 6722 | 61.45 | 57.49 | 64.89 | 51.48 | 0.953 | 0.855 | 1.047 | 0.982 |
| Inferior frontal gyrus, pars triangularis | 1615 | 12.97 | 10.79 | 15.66 | 12.72 | 0.139 | 0.114 | 0.170 | 0.151 |
| Inferior frontal gyrus, pars opercularis | 1877 | 29.33 | 25.73 | 33.49 | 26.63 | 0.347 | 0.297 | 0.408 | 0.337 |
| Precentral gyrus | 11055 | 68.34 | 65.09 | 71.40 | 55.33 | 1.150 | 1.052 | 1.252 | 1.172 |
| Temporal pole | 3341 | 6.21 | 4.56 | 8.37 | 4.14 | 0.064 | 0.047 | 0.087 | 0.056 |
| Superior temporal gyrus (A) | 622 | 1.11 | 0.68 | 1.84 | 0.59 | 0.011 | 0.007 | 0.019 | 0.006 |
| Superior temporal gyrus (P) | 2093 | 6.76 | 5.06 | 8.72 | 5.62 | 0.070 | 0.052 | 0.091 | 0.074 |
| Middle temporal gyrus (A) | 762 | 0.86 | 0.44 | 1.61 | 0.59 | 0.009 | 0.004 | 0.016 | 0.009 |
| Middle temporal gyrus (P) | 2794 | 7.04 | 5.29 | 9.04 | 6.51 | 0.073 | 0.054 | 0.095 | 0.077 |
| Middle temporal gyrus, temporooccipital part | 2470 | 10.65 | 8.11 | 13.47 | 9.17 | 0.113 | 0.085 | 0.145 | 0.109 |
| Inferior temporal gyrus (A) | 789 | 0.89 | 0.43 | 1.61 | 1.48 | 0.009 | 0.004 | 0.016 | 0.015 |
| Inferior temporal gyrus (P) | 2687 | 4.38 | 3.09 | 6.18 | 2.96 | 0.045 | 0.031 | 0.064 | 0.041 |
| Inferior temporal gyrus, temporoocipital part | 1883 | 11.95 | 9.60 | 14.69 | 11.24 | 0.127 | 0.101 | 0.159 | 0.139 |
| Postcentral gyrus | 7553 | 19.37 | 16.25 | 22.68 | 13.91 | 0.215 | 0.177 | 0.257 | 0.210 |
| Superior parietal lobule | 3626 | 29.07 | 25.56 | 32.94 | 28.99 | 0.343 | 0.295 | 0.400 | 0.358 |
| Supramarginal gyrus, (A) | 1951 | 10.43 | 8.25 | 12.98 | 9.76 | 0.110 | 0.086 | 0.139 | 0.112 |
| Supramarginal gyrus, (P) | 3017 | 26.64 | 23.05 | 30.39 | 26.92 | 0.310 | 0.262 | 0.362 | 0.346 |
| Angular gyrus | 2944 | 26.03 | 22.79 | 29.84 | 22.49 | 0.301 | 0.259 | 0.354 | 0.290 |
| Lateral occipital cortex (S) | 9535 | 68.22 | 64.60 | 71.56 | 54.73 | 1.146 | 1.039 | 1.257 | 1.145 |
| Lateral occipital cortex (I) | 5004 | 25.77 | 22.04 | 29.63 | 20.41 | 0.298 | 0.249 | 0.351 | 0.314 |
| Intracalcarine cortex | 2211 | 8.35 | 6.41 | 10.62 | 8.88 | 0.087 | 0.066 | 0.112 | 0.101 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Frontal medial cortex | 1473 | 1.13 | 0.61 | 1.89 | 1.18 | 0.011 | 0.006 | 0.019 | 0.012 |
| Juxtapositional lobule cortex | 2193 | 31.78 | 27.80 | 36.38 | 30.18 | 0.382 | 0.326 | 0.452 | 0.391 |
| Subcallosal cortex | 2150 | 2.55 | 1.59 | 3.99 | 1.48 | 0.026 | 0.016 | 0.041 | 0.021 |
| Paracingulate gyrus | 4095 | 38.76 | 34.81 | 42.96 | 41.12 | 0.490 | 0.428 | 0.561 | 0.524 |
| Cingulate gyrus (A) | 4144 | 19.72 | 16.63 | 23.34 | 18.64 | 0.220 | 0.182 | 0.266 | 0.240 |
| Cingulate gyrus (P) | 4668 | 10.78 | 8.30 | 13.93 | 10.65 | 0.114 | 0.087 | 0.150 | 0.139 |
| Precuneous cortex | 7439 | 27.51 | 23.96 | 31.35 | 21.89 | 0.322 | 0.274 | 0.376 | 0.308 |
| Cuneal cortex | 1605 | 6.87 | 5.19 | 8.92 | 6.21 | 0.071 | 0.053 | 0.093 | 0.065 |
| Frontal orbital cortex | 4058 | 20.06 | 16.85 | 23.39 | 15.98 | 0.224 | 0.185 | 0.266 | 0.222 |
| Parahippocampal gyrus (A) | 3225 | 3.74 | 2.51 | 5.36 | 2.66 | 0.038 | 0.025 | 0.055 | 0.041 |
| Parahippocampal gyrus (P) | 2014 | 4.22 | 3.02 | 5.89 | 3.55 | 0.043 | 0.031 | 0.061 | 0.041 |
| Lingual gyrus | 5387 | 15.50 | 12.74 | 18.71 | 13.31 | 0.168 | 0.136 | 0.207 | 0.160 |
| Temporal fusiform cortex (A) | 830 | 0.87 | 0.48 | 1.60 | 0.89 | 0.009 | 0.005 | 0.016 | 0.012 |
| Temporal fusiform cortex (P) | 2728 | 4.55 | 3.27 | 6.09 | 3.25 | 0.047 | 0.033 | 0.063 | 0.036 |
| Temporal occipital fusiform cortex | 2451 | 13.68 | 11.08 | 16.69 | 10.95 | 0.147 | 0.117 | 0.183 | 0.136 |
| Occipital fusiform gyrus | 3576 | 20.13 | 17.26 | 23.72 | 15.98 | 0.225 | 0.190 | 0.271 | 0.225 |
| Frontal operculum cortex | 1059 | 15.83 | 13.44 | 18.52 | 12.72 | 0.172 | 0.144 | 0.205 | 0.148 |
| Central operculum cortex | 2568 | 11.05 | 8.91 | 13.37 | 10.36 | 0.117 | 0.093 | 0.143 | 0.115 |
| Parietal operculum cortex | 1683 | 4.73 | 3.44 | 6.51 | 3.55 | 0.048 | 0.035 | 0.067 | 0.047 |
| Planum polare | 1210 | 2.45 | 1.71 | 3.42 | 1.78 | 0.025 | 0.017 | 0.035 | 0.021 |
| Heschl's gyrus | 786 | 2.22 | 1.43 | 3.29 | 2.07 | 0.022 | 0.014 | 0.033 | 0.021 |
| Planum temporale | 1434 | 4.37 | 3.13 | 5.77 | 4.73 | 0.045 | 0.032 | 0.059 | 0.050 |
| Supracalcarine cortex | 424 | 1.20 | 0.82 | 1.77 | 1.48 | 0.012 | 0.008 | 0.018 | 0.015 |
| Occipital pole | 4081 | 17.34 | 14.09 | 20.92 | 13.31 | 0.190 | 0.152 | 0.235 | 0.186 |

## A.5 Model assessment

We use posterior predictive checks [Gelman *et al.*, 1996] to assess how well our model fits the data. First, we generate 1,000 realisations $\mathbf{y}^*_{kt}$ from the posterior predictive distributions $\mathbf{Y}^*_{kt} \sim \mathcal{PP}\left(\mathcal{B}, \boldsymbol{\lambda}^{kt}\right)$, where $k = 1$ for, $k = 2$ for executive control, and $\boldsymbol{\lambda}^{kt}$ is the $t$-th draw from the posterior intensity of type $k$. We then use these observations to construct several summary statistics which we compare with the values observed in the real data. Both first and second order properties of the point patterns are considered.

For first order properties we study $N_{\mathbf{Y}^*_{kt}}(B)$, the total number of points in some region $B$ of the brain. For the entire brain, the posterior predictive samples yield 95% credible intervals of $[2, 13]$ for emotion and $[6, 20]$ for executively control. These cover 73% (623/855) and 63% (213/338) studies of the sample respectively. We further perform the same check for the 65 ROIs used in the full brain analysis of Section A.4. For each ROI and type, we record the percentage of studies that are contained in the 95% credible intervals of the posterior predictive distribution. Results are shown in Figure A.9. We also find that from the 1193 studies of the sample, 1148 (96%) have at least 90% of their ROI counts covered by the posterior predictive intervals.

Our second test is based on the $L$-function [Møller and Waagepetersen, 2004, 2007] which is a summary statistic for the second order properties of a point pattern. For our model, given an observation $\mathbf{y}^*_{kt}$ and an intensity estimate $\boldsymbol{\lambda}^{kt}$ the $L$-fucntion at distance $d$ can be estimated as:

$$L\left(d \mid \mathbf{y}^*_{kt}, \lambda^{kt}\right) = \left[\frac{3}{4\pi|\mathcal{B}|} \sum_{y_1, y_2 \in \mathbf{y}^*_{kt}} \frac{\mathbf{1}_{\{||y_1, y_2|| \leq d\}}}{\boldsymbol{\lambda}^{kt}_{\boldsymbol{\nu}_{y_1}}, \boldsymbol{\lambda}^{kt}_{\boldsymbol{\nu}_{y_2}}}\right]^{\frac{1}{3}}. \tag{A.15}$$

Therefore, as suggested by Illian *et al.* [2009], one can consider the differences $\Delta_{ik}(d) = L(d \mid \mathbf{x}_{ik}, \boldsymbol{\lambda}^{kt}) - L(d \mid \mathbf{y}^*_{kt}, \boldsymbol{\lambda}^{kt})$ for all $t = 1, \ldots, 1,000$, where $\mathbf{x}_{ik}$ is the $i$-th observation of type $k$. If the 95% credible interval of $\Delta_{ik}$ for some study $i$ does not contain zero then the model does a poor job explaining the data of this study. We investigate the differences $\Delta_{ik}(r)$ using values of $r$ in a grid ranging from 0mm to 200mm, with spacing 2mm. Figure A.10 summarises the results. In the top panel we see that the median (over studies) upper and lower bounds contain zero for both emotion and executive control studies. For each $r$, we calculate the proportion of studies that contain zero in the posterior predictive folders of $\Delta_{ik}$ and show results in the bottom panel of Figure A.10. We find that for $r \leq 40mm$ over

85% of the studies of both types contain zero within the credible interval. For larger distances, the proportion drops but remains well above 60% suggesting no major issues with the fit of our model.

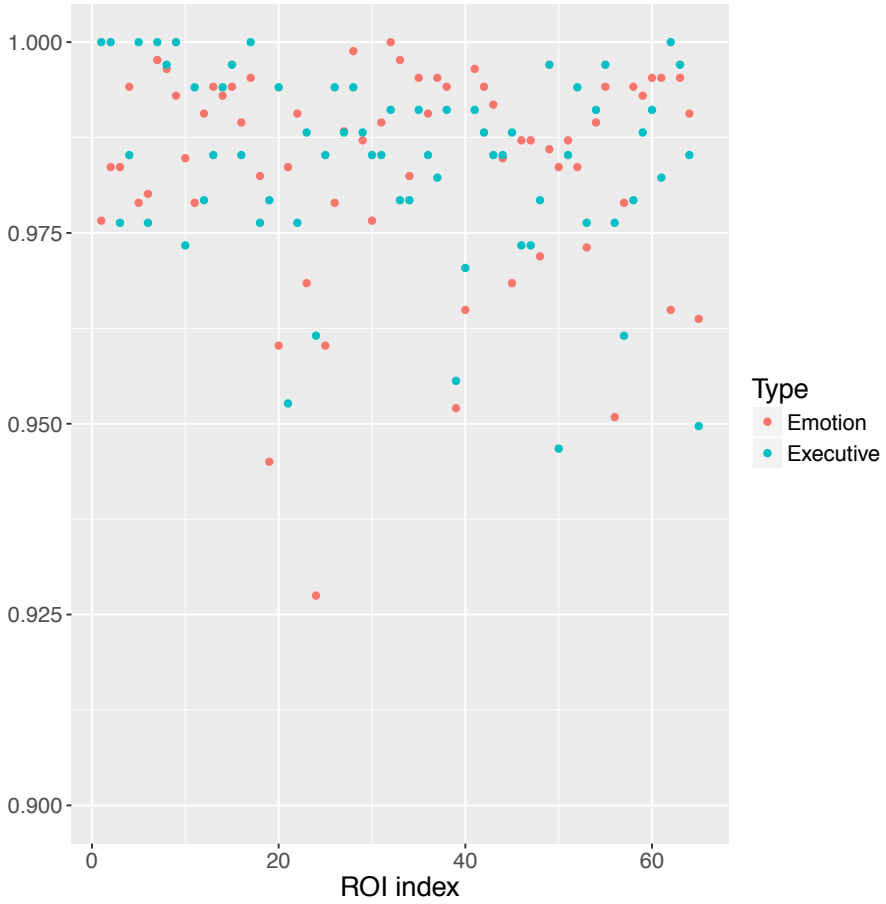



Figure A.9: Results of the first posterior predictive check for first order propoerties. The $x$-axis is the ROI index in the same order presented in the full brain analysis of Section A.4. The $y$-axis represents the percentage of our meta-analysis studies that have counts which fall within the 95% credible interval, as obtained from the 1,000 posterior predictive samples. Emotion studies are shown in red whereas executive control studies are shown in blue. The posterior predictive samples were generated from 1,000 intensity draws obtained from the HMC.

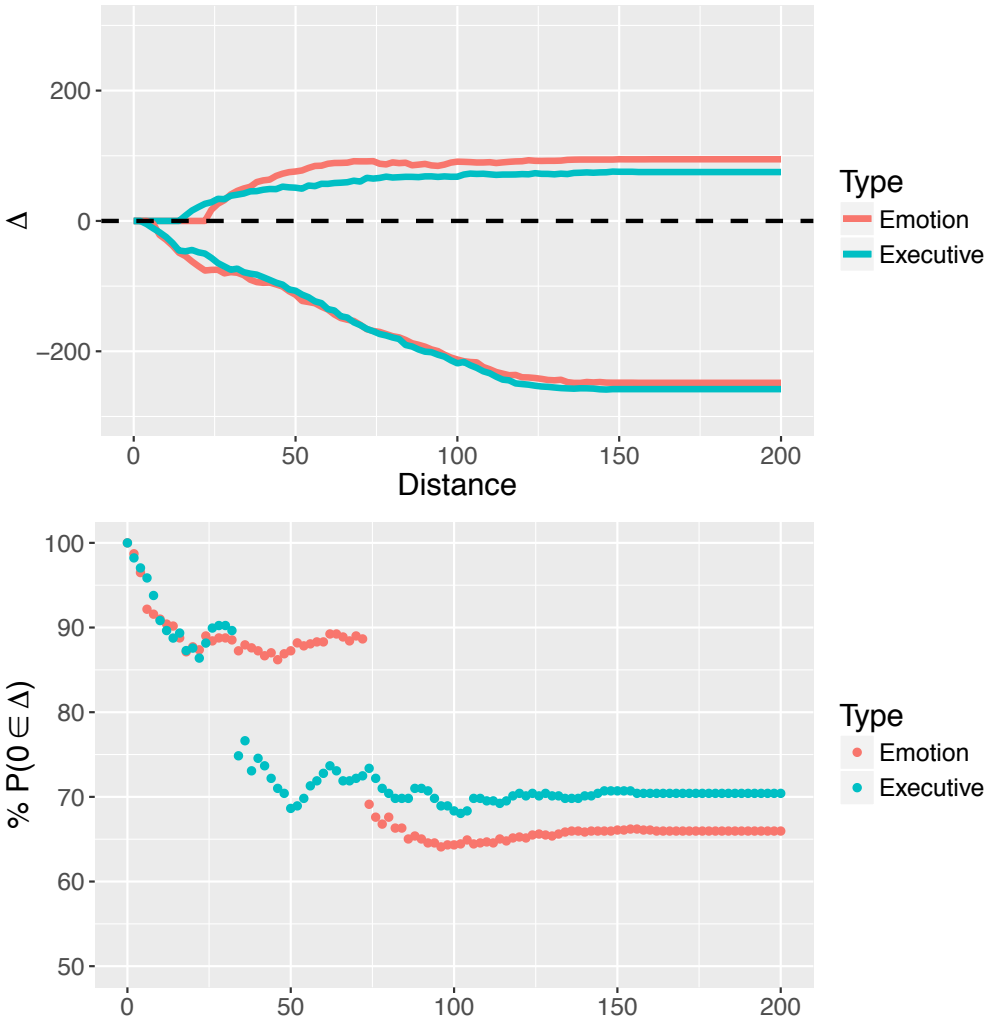



Figure A.10: Results of the second posterior predictive check for second order propoerties. The upper panel shows the median over studies upper and lower credible intervals for $\Delta_{ik}$. The bottom panel shows the proportion of studies which have credible intervals for $\Delta_{ik}$ that contain zero. Emotion studies are shown in red whereas executive control studies are shown in blue. The posterior predictive samples were generated from 1,000 intensity draws obtained from the HMC.

# Appendix B

# Group fMRI supplements

## B.1 Posterior distribution

For $i = 1, \ldots, I$ and $v = 1, \ldots, V$, let $\delta_{iv}$ be the allocation variables indicating which component each voxel belongs. From Equation (5.2) we *a priori* have that for all $i = 1, \ldots, I$ and $v = 1, \ldots, V$:

$$\pi\left(\delta_{iv} \mid m, \{x_{ij}, \mathbf{S}_{ij}\}_{j=1}^{n(\mathbf{x}_i)}\right) = \mathbb{P}\left(\delta_{iv} = j\right) = p_{ivj}. \tag{B.1}$$

We can write that:

$$\pi\left(y_{iv} \mid p_{iv0}, m_0, s_0^2, \left\{p_{ijv}, m_{ij}, s_{ij}^2\right\}_{j=1}^{n(\mathbf{x}_i)}, \delta_{iv}\right) = \\ \phi\left(y_{iv} \mid m_0, s_0^2\right)^{\mathbf{1}_{\{\delta_{iv}=0\}}} \prod_{j=1}^{n(\mathbf{x}_i)} \left[\phi\left(y_{iv} \mid m_{ij}, s_{ij}^2\right)^{\mathbf{1}_{\{\delta_{iv}=j\}}}\right]. \tag{B.2}$$

The posterior distribution of the model parameters, including allocation variables, conditional on the data and the fixed prior parameters $m$, $\mathbf{T}_\mathbf{x}$, $\epsilon$, $\eta$, $\mathbf{T}_0$, $\boldsymbol{\lambda}^+$ and $\boldsymbol{\lambda}^-$, is given up to a multiplicative constant as:

$$\pi\Bigg( \left\{\{\delta_{iv}\}_{v=1}^{V}, \left\{x_{ij}, \mathbf{S}_{ij}, m_{ij}, s_{ij}^2\right\}_{j=1}^{n(\mathbf{x}_i)}\right\}_{i=1}^{I}, m_0, s_0^2, \mu^+, \sigma_{m^+}^2, \beta_{s_+^2}, \mu^-, \sigma_{m^-}^2, \beta_{s_-^2}, \\ \left\{z_k, \boldsymbol{\Sigma}_k\right\}_{k=1}^{n(\mathbf{z})} \mathbf{T}_\mathbf{z} \mid \left\{\{y_{iv}\}_{v=1}^{V}\right\}_{i=1}^{I}, m, \mathbf{T}_\mathbf{x}, \epsilon, \eta, \mathbf{T}_0, \boldsymbol{\lambda}^+, \boldsymbol{\lambda}^- \Bigg) \propto$$

$$\prod_{i=1}^{I}\prod_{v=1}^{V}\left[\pi\left(y_{iv}\mid p_{iv0},m_0,s_0^2,\left\{p_{ijv},m_{ij},s_{ij}^2\right\}_{j=1}^{n(\mathbf{x}_i)},\delta_{iv}\right)\pi\left(\delta_{iv}\mid m,\left\{x_{ij},\mathbf{S}_{ij}\right\}_{j=1}^{n(\mathbf{x}_i)}\right)\right]$$
$$\prod_{i=1}^{I}\left[\pi\Big(\left\{x_{ij},\mathbf{S}_{ij},m_{ij},s_{ij}\right\}_{j=1}^{n(\mathbf{x}_i^+)}\mid\epsilon,\eta,\left\{z_k,\mathbf{\Sigma}_{z_k}\right\}_{k=1}^{n(\mathbf{z}^+)},\mathbf{T}_{\mathbf{x}},d_{\mathbf{x}},\mu^+,\sigma_{m^+}^2,\beta_{s_+^2}\Big)\right.$$
$$\left.\pi\Big(\left\{x_{ij},\mathbf{S}_{ij},m_{ij},s_{ij}\right\}_{j=n(\mathbf{x}_i^+)+1}^{n(\mathbf{x}_i^+)+n(\mathbf{x}_i^-)}\mid\epsilon,\eta,\left\{z_k,\mathbf{\Sigma}_{z_k}\right\}_{k=n(\mathbf{z}^+)+1}^{n(\mathbf{z}^+)+n(\mathbf{z}^-)},\mathbf{T}_{\mathbf{x}},d_{\mathbf{x}},\mu^-,\sigma_{m^-}^2,\beta_{s_-^2}\Big)\right]$$
$$\pi\left(m_0\right)\pi\left(s_0^2\right)\pi\left(\mu^+\right)\pi\left(\sigma_{m^+}^2\right)\pi\left(\beta_{s_+^2}\right)\pi\left(\mu^-\right)\pi\left(\sigma_{m^-}^2\right)\pi\left(\beta_{s_-^2}\right)$$
$$\pi\Big(\left\{z_k,\mathbf{\Sigma}_{z_k}\right\}_{k=1}^{n(\mathbf{z}^+)}\mid\lambda^+,\mathbf{T}_{\mathbf{z}},d_{\mathbf{z}}\Big)\pi\Big(\left\{z_k,\mathbf{\Sigma}_{z_k}\right\}_{k=n(\mathbf{z}^+)+1}^{n(\mathbf{z}^+)+n(\mathbf{z}^-)}\mid\lambda^-,\mathbf{T}_{\mathbf{z}},d_{\mathbf{z}}\Big)\pi\left(\mathbf{T}_{\mathbf{z}}\mid\mathbf{T}_0,d_0\right), \tag{B.3}$$

where $\pi\left(y_{iv}\mid p_{iv0},m_0,s_0^2,\left\{p_{ijv},m_{ij},s_{ij}^2\right\}_{j=1}^{n(\mathbf{x}_i)},\delta_{iv}\right)$ is defined in Equation (B.2), $\pi\Big(\delta_{iv}\mid m,\left\{x_{ij},\mathbf{S}_{ij}\right\}_{j=1}^{n(\mathbf{x}_i)}\Big)$ is defined in Equation (B.1) and all the remaining densities are defined in Section 5.2. The posterior distribution of Equation (B.3) cannot be solved analytically and hence we need to resort to MCMC to draw samples from it. We now present the details of the algorithm.

## B.2 Sampling algorithm details

In this section we provide the details of the MCMC algorithm used to sample from the posterior of Equation (B.3). We use a hybrid Gibbs algorithm where each parameter (or set of parameters) is drawn from its full conditional distribution. In particular, given initial values for the model parameters, we repeat the following steps for a pre-specified number of iterations (the type of update is indicated within parentheses):

1. Update $\delta_{iv}$ for all $v=1,\ldots,V$ and $i=1,\ldots,I$ (Gibbs)
2. Update $s_0^2$ (Gibbs)
3. Update $m_0$ (Gibbs)
4. Update $s_{ij}^2$ for all $i=1,\ldots,I$ and $j=1,\ldots,n\left(\mathbf{x}_i\right)$ (Gibbs)
5. Update $\beta_{s_+^2}$, $\beta_{s_-^2}$ (Gibbs)
6. Update $m_{ij}$ for all $i=1,\ldots,I$ and $j=1,\ldots,n\left(\mathbf{x}_i\right)$ (Gibbs)
7. Update $\mu^+$, $\mu^-$ (Gibbs)

8. Update $\sigma^2_{m^+}$, $\sigma^2_{m^-}$ (Gibbs)
9. Update $\{\mathbf{S}_{ij}\}_{j=1}^{n(\mathbf{x}_i)}$ for all $i = 1, \ldots, I$ (independence sampler)
10. Update $\{x_{ij}\}_{j=1}^{n(\mathbf{x}_i)}$ for all $i = 1, \ldots, I$ (independence sampler)
11. Update $n(\mathbf{x}_i^+)$, $n(\mathbf{x}_i^-)$ for all $i = 1, \ldots, I$ (spatial birth-and-death)
12. Update $\{z_k, \mathbf{\Sigma}_{z_k}\}_{k=1}^{n(\mathbf{z})}$ (independence sampler)
13. Update $n(\mathbf{z}^+)$, $n(\mathbf{z}^-)$ (spatial birth-and-death)
14. Update $\mathbf{T}_\mathbf{z}$ (Gibbs)

Initial values are drawn from the prior distributions. For individual and study centers we start with empty point configurations. Details for all the updates are given below. From now on, we denote $\pi(\boldsymbol{\theta} \mid \text{rest})$ the full conditional posterior distribution of parameter $\boldsymbol{\theta}$ given the remaining parameters of our model. For convenience, for $i = 1, \ldots, I$ we will sometimes use $m_{i0}$ to denote $m_0$ and $s^2_{i0}$ to denote $s^2_0$.

### Update $\delta_{iv}$ for all $v = 1, \ldots, V$ and $i = 1, \ldots, I$

We have that:

$$\pi(\delta_{iv} \mid \text{rest}) \propto \pi\left(y_{iv} \mid \{p_{ijv}, m_{ij}, s^2_{ij}\}_{j=0}^{n(\mathbf{x}_i)}, \delta_{iv}\right) \pi\left(\delta_{iv} \mid \{p_{ijv}\}_{j=0}^{n(\mathbf{x}_i)}\right). \tag{B.4}$$

Hence, we draw $\delta_{iv}$ from:

$$\mathbb{P}(\delta_{iv} = k) \propto \pi\left(y_{iv} \mid m_{ik}, s^2_{ik}, \delta_{iv} = k\right) p_{ivj}. \tag{B.5}$$

Once the allocation variables have been sampled, the expression of the likelihood in Equation B.2 simplifies. In particular, for each subject $i$, $i = 1, \ldots, I$ it is possible to write:

$$\prod_{v=1}^{V} \left[\pi\left(y_{iv} \mid p_{iv0}, m_0, s^2_0, \{p_{ijv}, m_{ij}, s^2_{ij}\}_{j=1}^{n(\mathbf{x}_i)}, \delta_{iv}\right) \pi\left(\delta_{iv} \mid \{p_{ijv}\}_{j=0}^{n(\mathbf{x}_i)}\right)\right] = \prod_{j=0}^{n(\mathbf{x}_i)} \prod_{v \in G_{ij}} \left[\pi\left(y_{iv} \mid m_{ij}, s^2_{ij}, \delta_{iv} = j\right) p_{ivj}\right], \tag{B.6}$$

where $G_{ij}$ is the set of voxels of participant $i$ for which $\delta_{iv} = j$. This simplification is used for the update of the remaining of our model parameters.

### Update $s_0^2$

The full conditional of $s_0^2$ is:

$$\pi\left(s_0^2 \mid \text{rest}\right) \propto \prod_{i=1}^{I} \prod_{v \in G_{i0}} \left[\left(s_0^2\right)^{-1/2} \exp\left(-\frac{\left(y_{iv} - m_0\right)^2}{2s_0^2}\right)\right] \left(s_0^2\right)^{-10^{-3}-1} \exp\left(-\frac{10^{-3}}{s_0^2}\right). \tag{B.7}$$

Hence, we draw

$$s_0^2 \sim \text{IG}(0.5 \sum_{i=1}^{i} n(G_{i0}) + 10^{-3}, 0.5 \sum_{i=1}^{I} \sum_{v \in G_{i0}} (y_{iv} - m_0)^2 + 10^{-3}).$$

### Update $m_0$

The full conditional of $m_0$ is:

$$\pi\left(m_0 \mid \text{rest}\right) \propto \prod_{i=1}^{I} \prod_{v \in G_{i0}} \left[\exp\left(-\frac{\left(y_{iv} - m_0\right)^2}{2s_0^2}\right)\right] \exp\left(-\frac{m_0^2}{2}\right). \tag{B.8}$$

Hence, we draw

$$m_0 \sim \mathcal{N}\left(\frac{s_0^{-2} \sum_{i=1}^{I} \sum_{v \in G_{i0}} y_{iv}}{s_0^{-2} \sum_{i=1}^{I} n\left(G_{i0}\right) + 1}, \frac{1}{s_0^{-2} \sum_{i=1}^{I} n\left(G_{i0}\right) + 1}\right).$$

### Update $s_{ij}^2$ for all $i = 1, \ldots, I$ and $j = 1, \ldots, n\left(\mathbf{x}_i\right)$

For any $i = 1, \ldots, I$ and $j = 1, \ldots, n\left(\mathbf{x}_i\right)$, the full conditional of $s_{ij}^2$ is:

$$\pi\left(s_{ij}^2 \mid \text{rest}\right) \propto \prod_{v \in G_{ij}} \left[\left(s_{ij}^2\right)^{-1/2} \exp\left(-\frac{\left(y_{iv} - m_{ij}\right)^2}{2s_{ij}^2}\right)\right] \left(s_{ij}^2\right)^{-2-1} \exp\left(-\frac{\beta_{s^2}}{s_{ij}^2}\right), \tag{B.9}$$

where $\beta_{s^2} = \beta_{s_+^2}$ for increases and $\beta_{s^2} = \beta_{s_-^2}$ for decreases. Hence, we draw

$$s_{ij}^2 \sim \text{IG}(0.5n(G_{ij}) + 2, 0.5 \sum_{v \in G_{ij}} (y_{iv} - m_{ij})^2 + \beta_{s^2}).$$

**Update $\beta_{s_+^2}$**

The full conditional of $\beta_{s_+^2}$ is:

$$\pi\left(\beta_{s_+^2} \mid \text{rest}\right) \propto \prod_{i=1}^{I} \prod_{j=1}^{n\left(\mathbf{x}_i^+\right)} \left[\beta_{s_+^2}^2 \exp\left(-\frac{\beta_{s_+^2}}{s_{ij}^2}\right)\right] \beta_{s_+^2}^{-10^{-2}-1} \exp\left(-10^{-2}\beta_{s_+^2}\right). \quad \text{(B.10)}$$

Hence we draw

$$\beta_{s_+^2} \sim \text{Ga}\left(2\sum_{i=1}^{I} n\left(\mathbf{x}_i^+\right) + 10^{-2}, \sum_{i=1}^{I} \sum_{j=1}^{n\left(\mathbf{x}_i^+\right)} s_{ij}^{-2} + 10^{-2}\right).$$

**Update $\beta_{s_-^2}$**

Similar to increases. We draw

$$\beta_{s_-^2} \sim \text{Ga}\Big(2\sum_{i=1}^{I} n\big(\mathbf{x}_i^-\big) + 10^{-2}, \sum_{i=1}^{I} \sum_{j=1+n\left(\mathbf{x}_i^+\right)}^{n\left(\mathbf{x}_i^+\right)+n\left(\mathbf{x}_i^-\right)} s_{ij}^{-2} + 10^{-2}\Big).$$

**Update $m_{ij}$ for all $i = 1, \ldots, I$ and $j = 1, \ldots, n\left(\mathbf{x}_i\right)$**

For any $i = 1, \ldots, I$ and $j = 1, \ldots, n\left(\mathbf{x}_i\right)$, the full conditional of $m_{ij}^2$ is:

$$\pi\left(m_{ij} \mid \text{rest}\right) \propto \prod_{v \in G_{ij}} \left[\exp\left(-\frac{\left(y_{iv} - m_{ij}\right)^2}{2s_{ij}^2}\right)\right] \exp\left(-\frac{\left(m_{ij} - \mu\right)^2}{2\sigma_m^2}\right) \mathbf{1}_{\{m_{ij} \in \mathcal{C}\}}, \quad \text{(B.11)}$$

where for increases $\left\{\mu, \sigma_m^2\right\} = \left\{\mu^+, \sigma_{m^+}^2\right\}$ and $\mathcal{C} = (0, \infty)$, whereas for decreases and $\left\{\mu, \sigma_m^2\right\} = \left\{\mu^-, \sigma_{m^-}^2\right\}$ and $\mathcal{C} = (-\infty, 0)$. Hence, we draw

$$m_{ij} \sim \mathcal{N}\left(AB^{-1}, B^{-1}\right),$$

where $A = \frac{\mu}{\sigma_m^2} + \frac{\sum_{v \in G_{ij}} y_{iv}}{s_{ij}^2}$ and $B = \frac{n(G_{ij})}{s_{ij}^2} + \frac{1}{\sigma_m^2}$, and accept if $m_{ij} \in \mathcal{C}$.

### Update $\mu^+$

The full conditional of $\mu^+$ is:

$$\pi\left(\mu^+ \mid \text{rest}\right) \propto \prod_{i=1}^{I} \prod_{j=1}^{n\left(\mathbf{x}_i^+\right)} \left[\exp\left(-\frac{\left(m_{ij}-\mu^+\right)^2}{2\sigma_{m^+}^2}\right)\right] \exp\left(-\frac{\left(\mu^+-3\right)^2}{20}\right). \tag{B.12}$$

Hence we draw

$$\mu^+ \sim \mathcal{N}\left(AB^{-1}, B^{-1}\right),$$

where $A = 0.3 + \frac{\sum_{i=1}^{I}\sum_{j=1}^{n\left(\mathbf{x}_i^+\right)} m_{ij}}{\sigma_{m^+}^2}$ and $B = 0.1 + \frac{\sum_{i=1}^{I} n\left(\mathbf{x}_i^+\right)}{\sigma_{m^+}^2}$.

### Update $\mu^-$

Similar to increases. We draw

$$\mu^- \sim \mathcal{N}\left(AB^{-1}, B^{-1}\right),$$

where $A = -0.3 + \frac{\sum_{i=1}^{I}\sum_{j=1+n\left(\mathbf{x}_i^+\right)}^{n\left(\mathbf{x}_i^+\right)+n\left(\mathbf{x}_i^-\right)} m_{ij}}{\sigma_{m^-}^2}$ and $B = 0.1 + \frac{\sum_{i=1}^{I} n\left(\mathbf{x}_i^-\right)}{\sigma_{m^-}^2}$.

### Update $\sigma_{m^+}^2$

The full conditional of $\sigma_{m^+}^2$ is:

$$\pi\left(\sigma_{m^+}^2 \mid \text{rest}\right) \propto \prod_{i=1}^{I} \prod_{j=1}^{n\left(\mathbf{x}_i^+\right)} \left[\left(\sigma_{m^+}^2\right)^{-1/2} \exp\left(-\frac{\left(m_{ij}-\mu^+\right)^2}{2\sigma_{m^+}^2}\right)\right] \left(\sigma_{m^+}^2\right)^{-10^{-2}-1} \exp\left(-\frac{10^{-2}}{\sigma_{m^+}^2}\right). \tag{B.13}$$

Hence we draw

$$\sigma_{m^+}^2 \sim \text{IG}\Big((0.5\sum_{i=1}^{I} n\left(\mathbf{x}_i^+\right) + 10^{-2}, 0.5\sum_{i=1}^{I}\sum_{j=1}^{n\left(\mathbf{x}_i^+\right)} \left(m_{ij}-\mu^+\right)^2 + 10^{-2}\Big).$$

### Update $\sigma^2_{m^-}$

Similar to increases. We draw

$$\sigma^2_{m^-} \sim \text{IG}\Big((0.5\sum_{i=1}^{I} n\left(\mathbf{x}_i^-\right) + 10^{-2}, 0.5\sum_{i=1}^{I}\sum_{j=1+n\left(\mathbf{x}_i^+\right)}^{n\left(\mathbf{x}_i^+\right)+n\left(\mathbf{x}_i^-\right)} \left(m_{ij} - \mu^-\right)^2 + 10^{-2}\Big).$$

### Update $\{\mathbf{S}_{ij}\}_{j=1}^{n(\mathbf{x}_i)}$ for all $i = 1, \ldots, I$ conditional on $n\left(\mathbf{x}_i\right)$

We update $\{\mathbf{S}_{ij}\}_{j=1}^{n(\mathbf{x}_i)}$ jointly using an independence sampler. For all $i = 1, \ldots, I$, the full conditional distribution is:

$$\pi\left(\{\mathbf{S}_{ij}\}_{j=1}^{n(\mathbf{x}_i)} \mid \text{rest}\right) \propto \frac{\prod_{j=1}^{n(\mathbf{x}_i)} \prod_{v\in G_{ij}} \left[\phi_3\left(\boldsymbol{\nu}_v; x_{ij}, \mathbf{S}_{ij}\right)\right]}{\prod_{v=1}^{V}\left[m + \sum_{k=1}^{n(\mathbf{x}_i)} \phi_3\left(\boldsymbol{\nu}_v; x_{ik}, \mathbf{S}_{ik}\right)\right]} \prod_{j=1}^{n(\mathbf{x}_i)} \pi\left(\mathbf{S}_{ij} \mid \mathbf{T}_{\mathbf{x}}, d_{\mathbf{x}}\right). \tag{B.14}$$

Hence, for $j = 1, \ldots, n(\mathbf{x}_i)$ we propose new values

$$\mathbf{S}^*_{ij} \sim \text{IW}\Big(\mathbf{T}_{\mathbf{x}} + \sum_{v\in G_{ij}} (\boldsymbol{\nu}_v - x_{ij})(\boldsymbol{\nu}_v - x_{ij})', d_{\mathbf{x}} + n(G_{ij})\Big)$$

and accept the new configuration with probability:

$$\min\left\{1, \prod_{v=1}^{V} \frac{m + \sum_{j=1}^{n(\mathbf{x})} \phi_3\left(\boldsymbol{\nu}_v; x_{ij}, \mathbf{S}_{ij}\right)}{m + \sum_{j=1}^{n(\mathbf{x})} \phi_3\left(\boldsymbol{\nu}_v; x_{ij}, \mathbf{S}^*_{ij}\right)}\right\}.$$

### Update $\{x_{ij}\}_{j=1}^{n(\mathbf{x}_i)}$ for all $i = 1, \ldots, I$ conditional on $n\left(\mathbf{x}_i\right)$

We update $\{x_{ij}\}_{j=1}^{n(\mathbf{x}_i)}$ jointly using an independence sampler. For all $i = 1, \ldots, I$, the full conditional distribution is:

$$\pi\left(\{x_{ij}\}_{j=1}^{n(\mathbf{x}_i)} \mid \text{rest}\right) \propto \frac{\prod_{j=1}^{n(\mathbf{x}_i)} \prod_{v\in G_{ij}} \left[\phi_3\left(\boldsymbol{\nu}_v; x_{ij}, \mathbf{S}_{ij}\right)\right]}{\prod_{v=1}^{V}\left[m + \sum_{k=1}^{n(\mathbf{x}_i)} \phi_3\left(\boldsymbol{\nu}_v; x_{ik}, \mathbf{S}_{ik}\right)\right]} \prod_{j=1}^{n(\mathbf{x}_i^+)} \Big[\epsilon + \eta \sum_{z_k\in\mathbf{z}^+} \phi_3\left(x_{ij}; z_k, \boldsymbol{\Sigma}_{z_k}\right)\Big] \prod_{j=n\left(\mathbf{x}_i^+\right)+1}^{n\left(\mathbf{x}_i^+\right)+n\left(\mathbf{x}_i^-\right)} \Big[\epsilon + \eta \sum_{z_k\in\mathbf{z}^-} \phi_3\left(x_{ij}; z_k, \boldsymbol{\Sigma}_{z_k}\right)\Big]. \tag{B.15}$$

Hence, for $j = 1, \ldots, n(\mathbf{x}_i)$ we propose new values $x^*_{ij} \sim \mathcal{N}_{\mathcal{B}}\Big(\frac{\sum_{v\in G_{ij}} \boldsymbol{\nu}_v}{n(G_{ij})}, \frac{\mathbf{S}_{ij}}{n(G_{ij})}\Big)$ and accept the new configuration with probability:

$$\min\Bigg\{1, \prod_{v=1}^{V}\Big[\frac{m+\sum_{j=1}^{n(\mathbf{x})}\phi_3\left(\boldsymbol{\nu}_v; x_{ij}, \mathbf{S}_{ij}\right)}{m+\sum_{j=1}^{n(\mathbf{x})}\phi_3\left(\boldsymbol{\nu}_v; x^*_{ij}, \mathbf{S}_{ij}\right)}\Big] \prod_{j=1}^{n(\mathbf{x}_i^+)}\Big[\frac{\epsilon+\eta\sum_{z_k\in\mathbf{z}^+}\phi_3\left(x^*_{ij}; z_k, \boldsymbol{\Sigma}_{z_k}\right)}{\epsilon+\eta\sum_{z_k\in\mathbf{z}^+}\phi_3\left(x_{ij}; z_k, \boldsymbol{\Sigma}_{z_k}\right)}\Big]$$
$$\prod_{j=n(\mathbf{x}_i^+)+1}^{n(\mathbf{x}_i^+)+n(\mathbf{x}_i^-)}\frac{\epsilon+\eta\sum_{z_k\in\mathbf{z}^-}\phi_3\left(x^*_{ij}; z_k, \boldsymbol{\Sigma}_{z_k}\right)}{\epsilon+\eta\sum_{z_k\in\mathbf{z}^-}\phi_3\left(x_{ij}; z_k, \boldsymbol{\Sigma}_{z_k}\right)}\Big]\Bigg\},$$

where $\mathcal{N}_{\mathcal{B}}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ is the three dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ truncated at the brain $\mathcal{B}$.

### Update $n(\mathbf{x}_i^+)$ for all $i = 1, \ldots, I$

We use the spatial birth-and-death algorithm to update the the total number of individual components for increases, $n(\mathbf{x}_i^+)$. Since we only want increase components to be placed in regions where the data $\{y_{iv}\}_{v=1}^{V}$ is positive, we use the following birth rate for a new point $\left\{x, \mathbf{S}, m, s^2\right\}$:

$$b\left(x, \mathbf{S}, m, s^2\right) = g^+\left(x\right)\pi(\mathbf{S} \mid \mathbf{T}_{\mathbf{x}}, d_{\mathbf{x}})\pi(m \mid \mu^+, \sigma^2_{m^+})\pi(s^2 \mid \beta_{s^2_+}), \tag{B.16}$$

with:

$$g^+\left(x\right) \propto c\sum_{v=1}^{V}\mathbf{1}_{\{x\in v\}}\mathbf{1}_{\{y_{iv}\geq 0\}}y_{iv}, \tag{B.17}$$

where $\{x \in v\}$ denotes the event that $x$ is inside the voxel indexed by $v$. Hence, for a birth we draw the location of an increase component from positive regions of the $T$ statistic image, with a bigger weights to voxels with high $y_{iv}$. The rest of the marks from their prior distributions. We choose $c$ such that $\int_{\mathcal{B}} g\left(\mathbf{x}\right) d\mathbf{x}$ is 1.

Then, the death rate is for removing a point $\left\{x_\ell, \mathbf{S}_\ell, m_\ell, s^2_\ell\right\}$, $1 \leq \ell \leq n\left(\mathbf{x}_i^+\right)$, from the current point configuration $\left\{x_{ij}, \mathbf{S}_{ij}, m_{ij}, s^2_{ij}\right\}_{j=1}^{n(\mathbf{x}_i^+)}$ can be derived from the detailed balance equation and is:

$$d\left(\left\{x_\ell, \mathbf{S}_\ell, m_\ell, s^2_\ell\right\}\right) = \prod_{v=1}^{V}\left[1+\frac{\phi_3\left(v \mid x_\ell, \mathbf{S}_\ell\right)}{m+\sum_{\substack{j=1\\ j\neq\ell}}^{n(\mathbf{x}_i^+)}\phi_3\left(v_{ij}, \mathbf{S}_{ij}\right)}\right]$$
$$\prod_{v\in G_{i\ell}}\left[\frac{p^*}{\phi_3\left(v \mid x_\ell, \mathbf{S}_\ell\right)}\frac{\phi\left(y_{iv} \mid m^*, s^{2*}\right)}{\phi\left(y_{iv} \mid m_\ell, s^2_\ell\right)}\right]\frac{g^+\left(x\ell\right)}{\rho_i^+\left(x_\ell\right)}, \tag{B.18}$$

where $p^* \in m \cup \{p_{ij}\}_{\substack{j=1 \\ j\neq\ell}}^{n(\mathbf{x}_i)-1}$, $m^* \in m_0 \cup \{m_{ij}\}_{\substack{j=1 \\ j\neq\ell}}^{n(\mathbf{x}_i)-1}$ and $s^{2*} \in s_0^2 \cup \left\{s_{ij}^2\right\}_{\substack{j=1 \\ j\neq\ell}}^{n(\mathbf{x}_i)-1}$ are the component parmeters under the new voxel allocation. We initialise $n\left(\mathbf{x}_i^+\right) = 0$ at the beginning of the simulation and then at each iteration set the total running time to $1/B = 1$.

### Update $n(\mathbf{x}_i^-)$ for all $i = 1, \ldots, I$

Analogous to increases. We use the spatial brith-and-death with birth rate for a new point $\left\{x, \mathbf{S}, m, s^2\right\}$:

$$b\left(x, \mathbf{S}, m, s^2\right) = g^-\left(x\right) \pi(\mathbf{S} \mid \mathbf{T}_{\mathbf{x}}, d_{\mathbf{x}}) \pi(m \mid \mu^-, \sigma_{m^-}^2) \pi(s^2 \mid \beta_{s_-^2}), \tag{B.19}$$

where:

$$g^-\left(x\right) \propto c \sum_{v=1}^{V} \mathbf{1}_{\{x \in v\}} \mathbf{1}_{\{y_{iv} \leq 0\}} y_{iv}. \tag{B.20}$$

We initialise $n\left(\mathbf{x}_i^-\right) = 0$ at the beginning of the simulation and set the total simulation time at $1/B = 1$ at each iteration.

### Update $\{z_k, \mathbf{\Sigma}_{z_k}\}_{k=1}^{n\left(\mathbf{z}^+\right)}$ conditional on $n\left(\mathbf{z}^+\right)$

Following Kang *et al.* [2011] we introduce the latent variables $\zeta_{x_{ij}}$, $x_{ij} \in \mathbf{x}_i^+$, that indicate to which increase study center each individual increase component is assigned. Components can of course be background an so $\zeta_{x_{ij}} \in \{\emptyset \cup \mathbf{z}^+\}$. We *a priori* assume that for all $z_k \in \{\emptyset \cup \mathbf{z}^+\}$, $\mathbb{P}\left(\zeta_{x_{ij}} = z_k\right) = \frac{1}{1+n(\mathbf{z}^+)}$. Given these latent variables the new intensities are:

$$\rho^+\left(x_{ij}\right) = \begin{cases} \epsilon & \zeta_{x_{ij}} = 0 \\ \phi_3\left(x_{ij} \mid z_k, \mathbf{\Sigma}_{z_k}\right) & \zeta_{x_{ij}} = z_k \end{cases}. \tag{B.21}$$

Let $c_{z_k}$ be the total number of increase components $x_{ij}$ for which $\zeta_{xij} = z_k$. For all $z_k \in \mathbf{z}^+$, the full conditional can be written as:

$$\pi\left(z_k \mid \text{rest}\right) \propto \boldsymbol{\lambda}^+\left(z_k\right) \exp\left(-c_{z_k} \eta \int_{\mathcal{B}} \phi_3\left(\xi \mid z_k, \mathbf{\Sigma}_{z_k}\right) d\xi\right) \phi_3\left(z_k \mid \frac{1}{c_{z_k}} \sum_{i=1}^{I} \sum_{j=1}^{n\left(\mathbf{x}_i^+\right)} x_{ij} \mathbf{1}_{\left\{\zeta_{x_{ij}} = z_k\right\}}, \frac{1}{c_{z_k}} \mathbf{\Sigma}_{z_k}\right). \tag{B.22}$$

Therefore we draw

$$z_k^* \sim \mathcal{N}_{\mathcal{B}} \left( \frac{1}{c_{z_k}} \sum_{i=1}^{I} \sum_{j=1}^{n(\mathbf{x}_i^+)} x_{ij} \mathbf{1}_{\{\zeta_{x_{ij}} = z_k\}}, \frac{1}{c_{z_k}} \mathbf{\Sigma}_{z_k} \right)$$

and accept with probability

$$\min \left\{ 1, \frac{\boldsymbol{\lambda}^+ (z_k^*)}{\boldsymbol{\lambda}^+ (z_k)} \exp \left( c_{z_k} \eta \left( \int_{\mathcal{B}} \phi \left( \xi \mid z_k, \mathbf{\Sigma}_{z_k} \right) d\xi - \int_{\mathcal{B}} \phi \left( \xi \mid z_k^*, \mathbf{\Sigma}_{z_k} \right) d\xi \right) \right) \right\}.$$

For the marks $\mathbf{\Sigma}_{z_k}$ we have that:

$$\pi \left( \mathbf{\Sigma}_{z_k} \mid \text{rest} \right) \propto \pi \left( \mathbf{\Sigma}_{z_k} \mid \mathbf{T}_{\mathbf{z}} \right) \exp \left( -c_{z_k} \eta \int_{\mathcal{B}} \phi_3 \left( \xi \mid z_k, \mathbf{\Sigma}_{z_k} \right) d\xi \right) \phi_3 \left( z_k \mid \frac{1}{c_{z_k}} \sum_{i=1}^{I} \sum_{j=1}^{n(\mathbf{x}_i^+)} x_{ij} \mathbf{1}_{\{\zeta_{x_{ij}} = z_k\}}, \frac{1}{c_{z_k}} \mathbf{\Sigma}_{z_k} \right). \quad \text{(B.23)}$$

Hence we draw:

$$\mathbf{\Sigma}_{z_k}^* \sim \text{IW} \left( \mathbf{T}_{\mathbf{z}} + \sum_{i=1}^{I} \sum_{j=1}^{n(\mathbf{x}_i^+)} \left( x_{ij} - z_k \right) \left( x_{ij} - z_k \right)^{\mathrm{T}} \mathbf{1}_{\{\zeta_{x_{ij}} = z_k\}}, d_{\mathbf{z}} + c_{z_k} \right)$$

and accept with probability

$$\min \left\{ 1, \exp \left( c_{z_k} \eta \left( \int_{\mathcal{B}} \phi \left( \xi \mid z_k, \mathbf{\Sigma}_{z_k} \right) d\xi - \int_{\mathcal{B}} \phi \left( \xi \mid z_k, \mathbf{\Sigma}_{z_k}^* \right) d\xi \right) \right) \right\}.$$

Finally for the latent parameters $\zeta_{x_{ij}}$ we *a posteriori* have that:

$$\mathbb{P} \left( \zeta_{x_{ij}} = \emptyset \right) = \frac{\epsilon}{\epsilon + \eta \sum_{z_\ell \in \mathbf{z}} \phi \left( x_{ij} \mid z_\ell, \mathbf{\Sigma}_{z_\ell} \right)}, \quad \text{(B.24)}$$

and:

$$\mathbb{P} \left( \zeta_{x_{ij}} = z_k \right) = \frac{\phi \left( x_{ij} \mid z_k, \mathbf{\Sigma}_{z_k} \right)}{\epsilon + \eta \sum_{z_\ell \in \mathbf{z}} \phi \left( x_{ij} \mid z_\ell, \mathbf{\Sigma}_{z_\ell} \right)}, \quad \text{(B.25)}$$

We use Equations (B.24) and (B.25) to sample $\zeta_{x_{ij}}$. The updates that we described require the calculation of normal probabilities over the brain. Following Kang *et al.* [2011] we evaluate those with a simple accept/reject Monte Carlo algorithm unless the 99% ellipsoids lie within the brain in which case we set them equal to 0.995.

**Update $\{z_k, \mathbf{\Sigma}_{z_k}\}_{k=n(\mathbf{z}^+)+1}^{n(\mathbf{z}^+)+n(\mathbf{z}^-)}$ conditional on $n(\mathbf{z}^-)$**

Similar to increases. For $z \in \mathbf{z}^-$ we draw

$$z_k^* \sim \mathcal{N}_{\mathcal{B}}\left(\frac{1}{c_{z_k}}\sum_{i=1}^{I}\sum_{j=n(\mathbf{x}_i^+)+1}^{n(\mathbf{x}_i^+)+n(\mathbf{x}_i^-)} x_{ij}\mathbf{1}_{\{\zeta_{x_{ij}}=z_k\}}, \frac{1}{c_{z_k}}\mathbf{\Sigma}_{z_k}\right)$$

and accept with probability

$$\min\left\{1, \frac{\boldsymbol{\lambda}^+(z_k^*)}{\boldsymbol{\lambda}^+(z_k)}\exp\left(c_{z_k}\eta\left(\int_{\mathcal{B}}\phi(\xi \mid z_k, \mathbf{\Sigma}_{z_k})\,d\xi - \int_{\mathcal{B}}\phi(\xi \mid z_k^*, \mathbf{\Sigma}_{z_k})\,d\xi\right)\right)\right\}.$$

For the marks $\mathbf{\Sigma}_{z_k}$, $z_k \in \mathbf{z}^-$ we draw:

$$\mathbf{\Sigma}_{z_k}^* \sim \mathrm{IW}\left(\mathbf{T}_{\mathbf{z}} + \sum_{i=1}^{I}\sum_{j=n(\mathbf{x}_i^+)+1}^{n(\mathbf{x}_i^+)+n(\mathbf{x}_i^-)}(x_{ij}-z_k)(x_{ij}-z_k)^{\mathrm{T}}\mathbf{1}_{\{\zeta_{x_{ij}}=z_k\}}, d_{\mathbf{z}}+c_{z_k}\right)$$

and accept with probability

$$\min\left\{1, \exp\left(c_{z_k}\eta\left(\int_{\mathcal{B}}\phi(\xi \mid z_k, \mathbf{\Sigma}_{z_k})\,d\xi - \int_{\mathcal{B}}\phi\left(\xi \mid z_k, \mathbf{\Sigma}_{z_k}^*\right)d\xi\right)\right)\right\}.$$

**Update $n(\mathbf{z}^+)$**

We update the total number of study centers using the spatial birth-and-death process. Since we want to place study centers in regions where individual components appear, we use the following birth rate for a new point $\{z, \mathbf{\Sigma}_z\}$:

$$b(z, \mathbf{\Sigma}_z) = \eta\sum_{i=1}^{I}\sum_{x_{ij}\in\mathbf{x}_i^+}\left[\phi_3(z \mid x_{ij}, \mathbf{\Sigma}_z)\,\pi(\mathbf{\Sigma}_z \mid \mathbf{T}_{\mathbf{z}}, d_{\mathbf{z}})\right]. \tag{B.26}$$

The total birth rate is then:

$$B = \eta\sum_{i=1}^{I}\sum_{x_{ij}\in\mathbf{x}_i^+}\left[\int_{\mathcal{B}}\int_{\mathcal{M}}\phi_3(z \mid x_{ij}, \mathbf{\Sigma}_z)\,\pi(\mathbf{\Sigma}_z \mid \mathbf{T}_{\mathbf{z}}, d_{\mathbf{z}})\,d\mathbf{\Sigma}_z dz\right], \tag{B.27}$$

where $\mathcal{M}$ is the space of $3\times 3$ positive definite matrices. The integral:

$$\int_{\mathcal{M}}\phi_3(z \mid x_{ij}, \mathbf{\Sigma}_z)\,\pi(\mathbf{\Sigma}_z \mid \mathbf{T}_{\mathbf{z}}, d_{\mathbf{z}})\,d\mathbf{\Sigma}_z$$

is the marginal distribution of $z$ in the Normal-Inverse Wishart model where $\mathbf{\Sigma}_z \mid \mathbf{T}_\mathbf{z}, d_\mathbf{z} \sim \mathrm{IW}\left(\mathbf{T}_\mathbf{z}, d_\mathbf{z}\right)$ and $z \mid x_{ij}, \mathbf{\Sigma}_z \sim \mathcal{N}_3\left(x_{ij}, \mathbf{\Sigma}_z\right)$. This marginal is known to be the three dimensional $t$ distribution with location $x_{ij}$, scale matrix $\frac{\mathbf{\Sigma}_z}{d_\mathbf{z}-2}$ and $d_\mathbf{z}-2$ degrees of freedom [Gelman *et al.*, 2013, p. 73]. As a result, the total birth rate can be written as:

$$B = \eta \sum_{i=1}^{I} \sum_{x_{ij} \in \mathbf{x}_i^+} \left[ \int_{\mathcal{B}} t_{3,d_\mathbf{z}-2}\left(z \mid x_{ij}, \frac{\mathbf{T}_\mathbf{z}}{d_\mathbf{z}-2}\right) dz \right] \tag{B.28}$$

where $t_{d,\kappa}\left(x \mid \mu, \mathbf{\Sigma}\right)$ is density function of the $d$-dimensional $t$ distribution with parameters $\mu, \mathbf{\Sigma}$ and $\kappa$ degrees of freedom, evaluated at $x$. For a birth we draw $\{z, \mathbf{\Sigma}_z\}$ from :

$$\begin{aligned} \frac{b\left(z, \mathbf{\Sigma}_z\right)}{B} &= \sum_{i=1}^{I} \sum_{x_{ij} \in \mathbf{x}_i^+} \left[ \frac{\phi_3\left(z \mid x_{ij}, \mathbf{\Sigma}_z\right) \pi\left(\mathbf{\Sigma}_z \mid \mathbf{T}_\mathbf{z}, d_\mathbf{z}\right)}{B} \right] \\ &= \sum_{i=1}^{I} \sum_{x_{ij} \in \mathbf{x}_i^+} \left[ \frac{\phi_3\left(z \mid x_{ij}, \mathbf{\Sigma}_z\right) \pi\left(\mathbf{\Sigma}_z \mid \mathbf{T}_\mathbf{z}, d_\mathbf{z}\right)}{\int_{\mathcal{B}} t_{3,d_\mathbf{z}-2}\left(z \mid x_{ij}, \frac{\mathbf{T}_\mathbf{z}}{d_\mathbf{z}-2}\right) dz} \frac{\int_{\mathcal{B}} t_{3,d_\mathbf{z}-2}\left(z \mid x_{ij}, \frac{\mathbf{T}_\mathbf{z}}{d_\mathbf{z}-2}\right) dz}{B} \right]. \end{aligned} \tag{B.29}$$

This formulation implies that we first draw $\mathbf{\Sigma}_z \sim \mathrm{IW}\left(\mathbf{T}_\mathbf{z}, d_\mathbf{z}\right)$ and then draw $z$ from the mixture $\sum_{i=1}^{I} \sum_{x_{ij} \in \mathbf{x}_i^+} \left[ w_{x_{ij}} \mathcal{N}_\mathcal{B}\left(x_{ij}, \mathbf{\Sigma}_z\right) \right]$ where the mixing wights $w_{x_{ij}} = \frac{1}{B} \int_{\mathcal{B}} t_{3,d_\mathbf{z}-2}\left(z \mid x_{ij}, \frac{\mathbf{T}_\mathbf{z}}{d_\mathbf{z}-2}\right) dz$.

The full conditional of the process $\{z_k, \mathbf{\Sigma}_{z_k}\}_{k=1}^{n\left(\mathbf{z}^+\right)}$ is:

$$\begin{aligned} \pi\left(\{z_k, \mathbf{\Sigma}_{z_k}\}_{k=1}^{n\left(\mathbf{z}^+\right)} \mid \text{rest}\right) \propto \prod_{i=1}^{I} \left[ \exp\left( -\epsilon|\mathcal{B}| - \int_{\mathcal{B}} \eta \sum_{z_k \in \mathbf{z}^+} \phi\left(\xi \mid z_k \mathbf{\Sigma}_{z_k} d\xi\right) \right) \right] \\ \prod_{i=1}^{I} \prod_{x_{ij} \in \mathbf{x}_i^+} \left[ \epsilon + \eta \sum_{z_k \in \mathbf{z}^+} \phi\left(x_{ij} \mid z_k, \mathbf{\Sigma}_{z_k}\right) \right] \prod_{z_k \in \mathbf{z}^+} \left[ \lambda^+\left(z_k\right) \pi\left(\mathbf{\Sigma}_{z_k} \mid \mathbf{T}_\mathbf{z}, d_\mathbf{z}\right) \right]. \end{aligned} \tag{B.30}$$

Hence, the death rate for removing a point $\{z_\ell, \mathbf{\Sigma}_{z_\ell}\}$, $1 \leq \ell \leq n\left(\mathbf{z}^+\right)$ from the current point configuration $\{z_k, \mathbf{\Sigma}_{z_k}\}_{k=1}^{n\left(\mathbf{z}^+\right)}$ is:

$$d\left(\{z_\ell, \mathbf{\Sigma}_{z_\ell}\}\right) = \frac{\eta \exp\left(\eta \int_{\mathcal{B}} \phi\left(\xi \mid z_\ell, \mathbf{\Sigma}_{z_\ell}\right)\right) \sum_{i=1}^{I} \sum_{x_{ij} \in \mathbf{x}_i^+} \left[ \phi_3\left(z_\ell \mid x_{ij}, \mathbf{\Sigma}_{z_\ell}\right) \right]}{\prod_{i=1}^{I} \prod_{x_{ij} \in \mathbf{x}_i^+} \left[ 1 + \frac{\eta \phi\left(x_{ij} | z_\ell, \mathbf{\Sigma}_{z_\ell}\right)}{\rho_i^+\left(x_{ij} | \cdot\right) - \eta \phi\left(x_{ij} | z_\ell, \mathbf{\Sigma}_{z_\ell}\right)} \right] \boldsymbol{\lambda}^+\left(z_\ell\right)}. \tag{B.31}$$

We initialise $n\left(z^{+}\right)=0$ and at each iteration set the total running time to $1/B$. The $t$ and normal integrals over the brain are approximated with an accept/reject Monte Carlo algorithm. We use 1,000 Monte Carlo iterations.

### Update $n\left(\mathbf{z}^{-}\right)$

Similar to increases. We use the spatial birth-and-death with birth rate for a new point $\{z, \mathbf{\Sigma}_z\}$:

$$b\left(z, \mathbf{\Sigma}_z\right)=\eta \sum_{i=1}^{I} \sum_{x_{ij} \in \mathbf{x}_i^{-}}\left[\phi_3\left(z \mid x_{ij}, \mathbf{\Sigma}_z\right) \pi\left(\mathbf{\Sigma}_z \mid \mathbf{T}_{\mathbf{z}}, d_{\mathbf{z}}\right)\right]. \tag{B.32}$$

We initialise $n\left(z^{+}\right)=0$ and at each iteration set the total running time to $1/B$.

### Update $\mathbf{T}_{\mathbf{z}}$

The full conditional of $\mathbf{T}_{\mathbf{z}}^{-1}$ is:

$$\pi\left(\mathbf{T}_{\mathbf{z}}^{-1} \mid \text{rest}\right) \propto\left|\mathbf{T}_{\mathbf{z}}\right|^{\frac{d_{\mathbf{z}} n(\mathbf{z})}{2}} \exp\left(-\frac{1}{2} \operatorname{tr}\left(\mathbf{T}_{\mathbf{z}} \sum_{z \in \mathbf{z}} \mathbf{\Sigma}_z^{-1}\right)\right)\left|\mathbf{T}_{\mathbf{z}}\right|^{\frac{d_0+p}{2}} \exp\left(-\frac{1}{2} \operatorname{tr}\left(\mathbf{T}_0 \mathbf{T}_{\mathbf{z}}\right)\right). \tag{B.33}$$

Hence we draw

$$\mathbf{T}_{\mathbf{z}}^{-1} \sim IW\left(\mathbf{T}_0+\sum_{z \in \mathbf{z}} \mathbf{\Sigma}_z^{-1}, n\left(\mathbf{z}\right) d_{\mathbf{z}}+d_0\right).$$

## B.3 Real data analysis supplementary figures

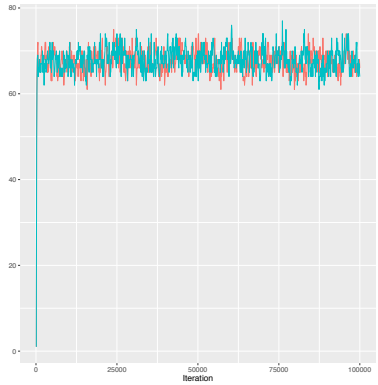
In this Section we provide supplementary plots to the real data analysis of Section 5.5. Figures B.1 and B.2 show the total number of increase $n\left(\mathbf{x}_i^{+}\right)$ and decrease components $n\left(\mathbf{x}_i^{-}\right)$ for individuals 1-3 and 4-6 respectively. Figure B.3 includes the total number of increase $n\left(\mathbf{z}^{+}\right)$ and decreases $n\left(\mathbf{z}^{-}\right)$ study centers, the background component parameters $m_0$ and $s_0^2$, and the variance hyperparameters $\beta_{s_+^2}$ and $\beta_{s_-^2}$. Finally we show the individual component hyperparameters $\mu^{+}$, $\mu^{-}$, $\sigma_{m^{+}}^2$ and $\sigma_{m^{-}}^2$ in Figure B.4.

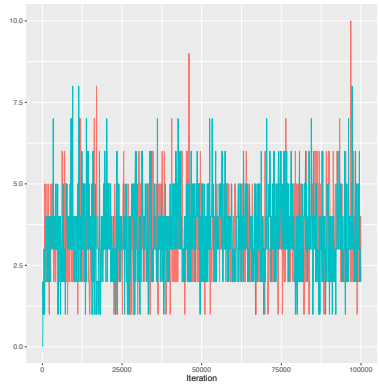
Figures B.5 and B.6 show marginal posterior probabilities of observing an increase study center at each voxel. The difference compared to our initial run, for which results were presented in Figure 5.10, is that for B.5 we have changed the expected number of background voxels to $q=95\%$ whereas for B.6 we have set

the expected number of components per center $\eta$ to 6, all other prior specifications remaining constant.

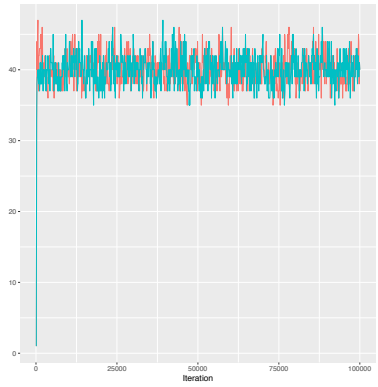
In Figures B.8, B.9, B.10, B.11 andB.12 we show posterior probabilities of activation and deactivation for participants 2-6 of the group analysis. Finally, in Figure B.7 we show the number of participants that had missing data at a given voxel, for several axial slices.

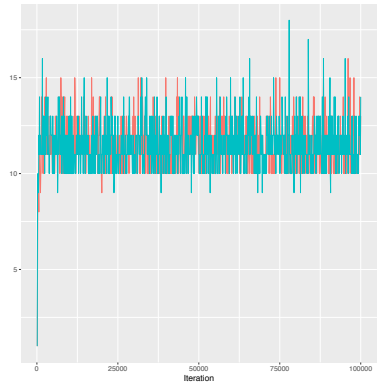
**Participant 1 increase components** **Participant 1 decrease components**



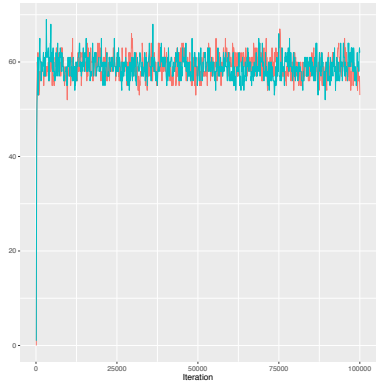




**Participant 2 increase components** **Participant 2 decrease components**




**Participant 3 increase components** **Participant 3 decrease components**



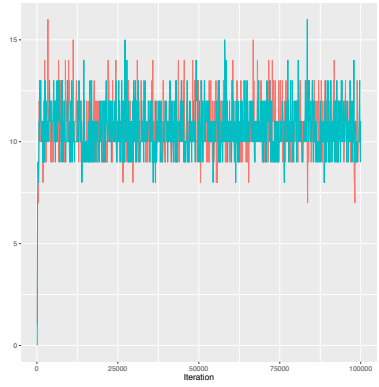




Figure B.1: Posterior traceplots for the group fMRI application of Section 5.5. The figure shows the total number of individual increase (left) and decrease components (right) for participants 1-3 of the experiment. All 100,000 posterior draws are presented.

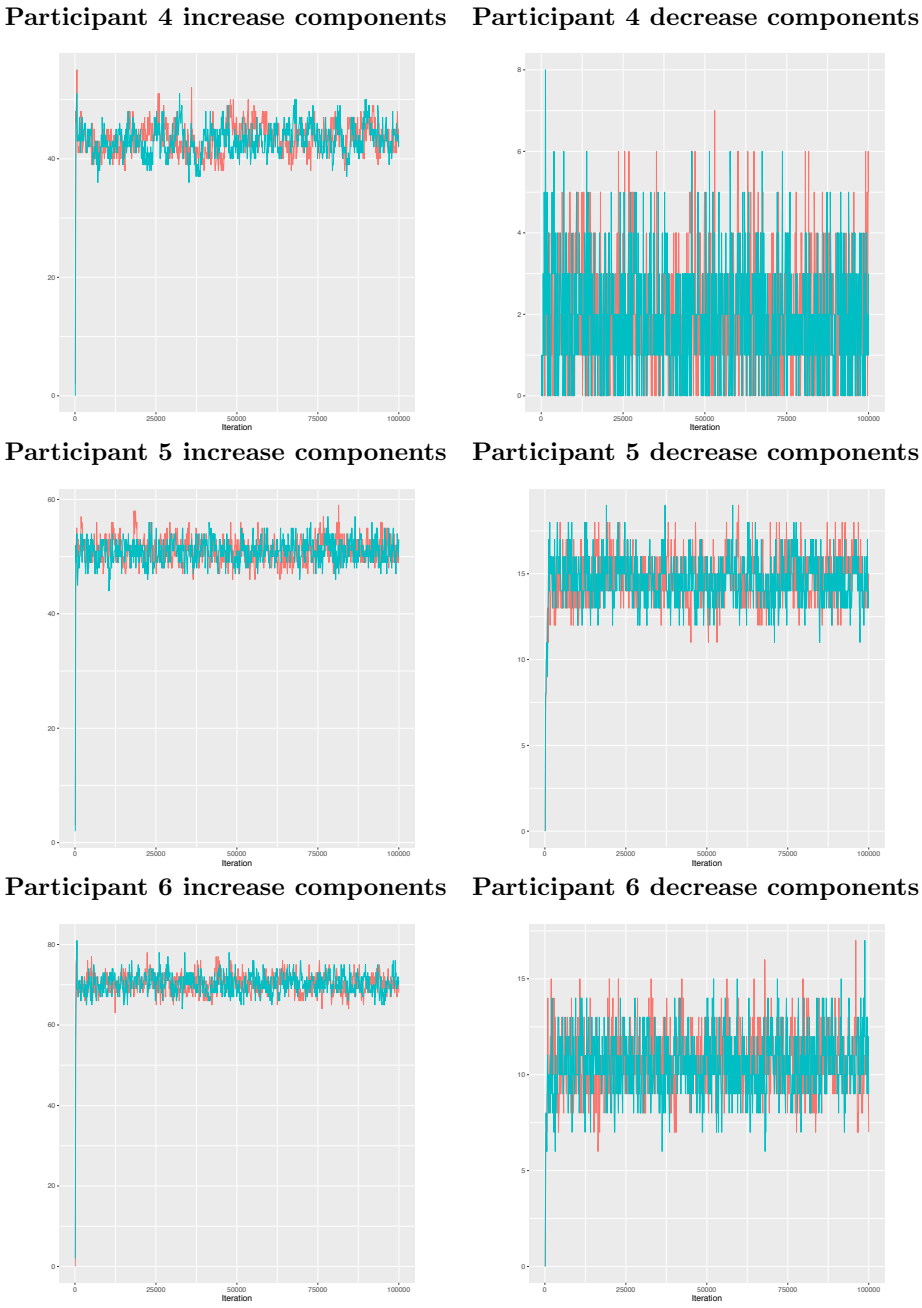



Figure B.2: Posterior traceplots for the group fMRI application of Section 5.5. The figure shows the total number of individual increase (left) and decrease components (right) for participants 4-6 of the experiment. All 100,000 posterior draws are presented.

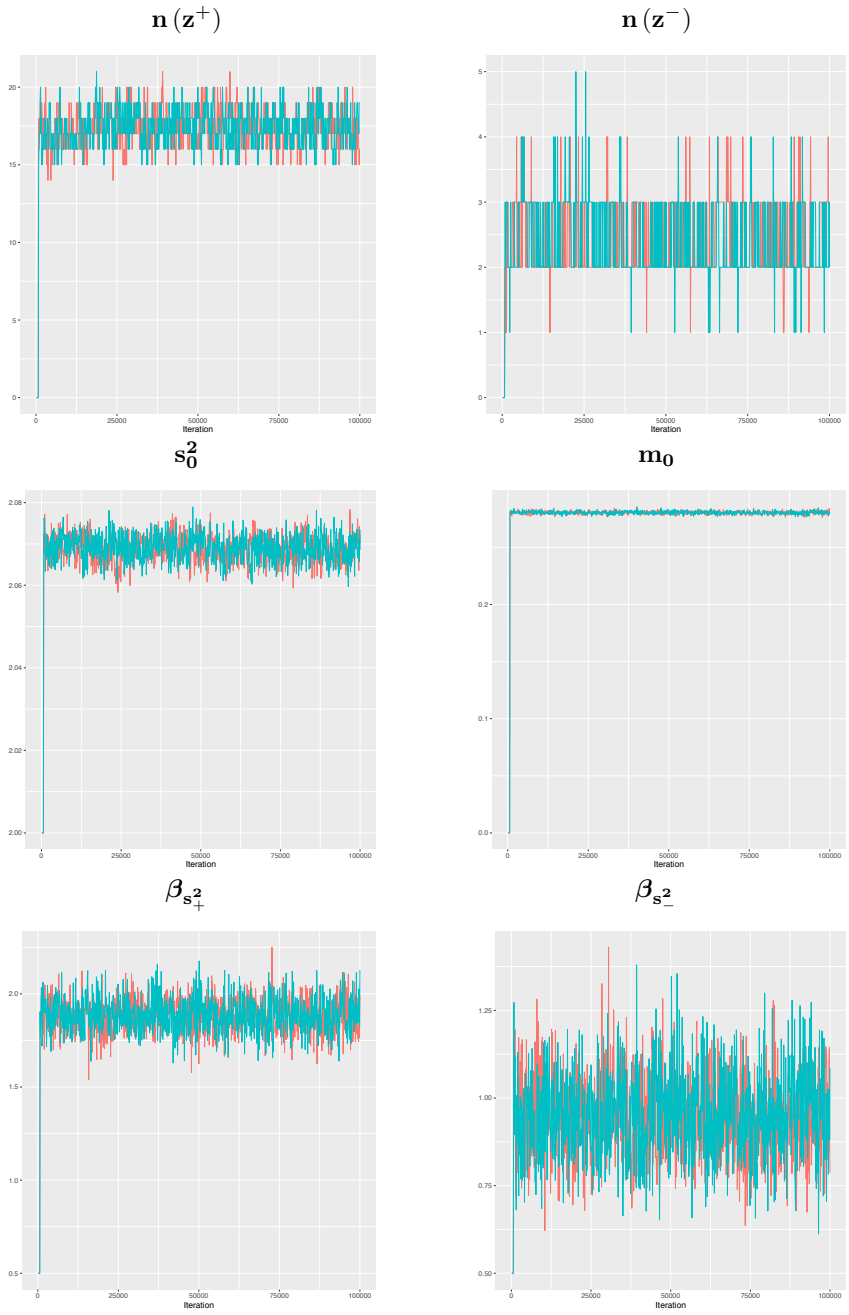



Figure B.3: Posterior traceplots for the group fMRI application of Section 5.5. The figure shows the total number of increase and decrease study centers (top row), the background component parameters $s_0^2$ and $m_0$ (middle row) and the variance hyper-parameters $\beta_{s_+^2}$ and $\beta_{s_-^2}$ (bottom row). All 100,000 posterior draws are presented.

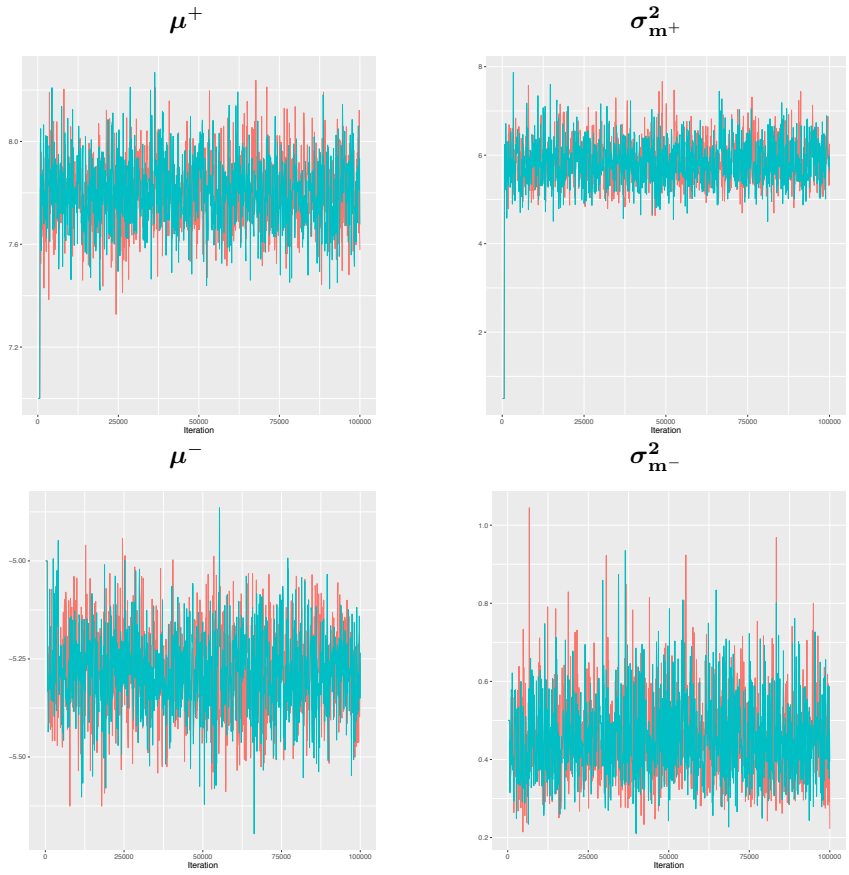



Figure B.4: Posterior traceplots for the group fMRI application of Section 5.5. The figure shows individual increase component hyperparameters $\mu^+$, $\sigma^2_{m^+}$ (top row) and individual component decreases hyperparameters $\mu^-$, $\sigma^2_{m^-}$ (bottom row). All 100,000 draws are presented. B.4.

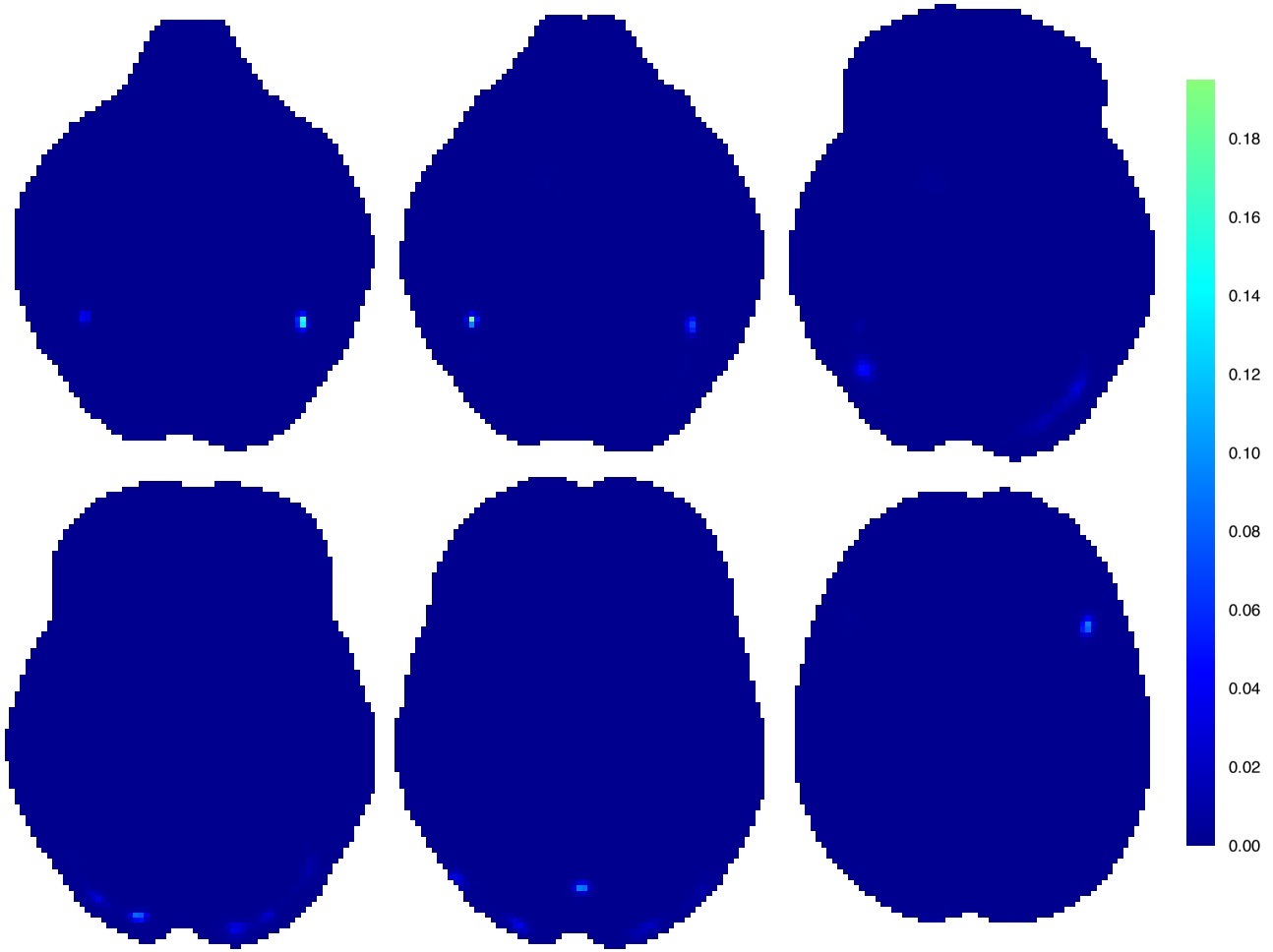



Figure B.5: Real data sensitivity analysis results. For each voxel, the figure shows the marginal posterior probabilities of observing an increase study center. In this run we use the same prior specifications defined in Section 5.5.2 except from $q$ which we set as 95% instead of 99%.

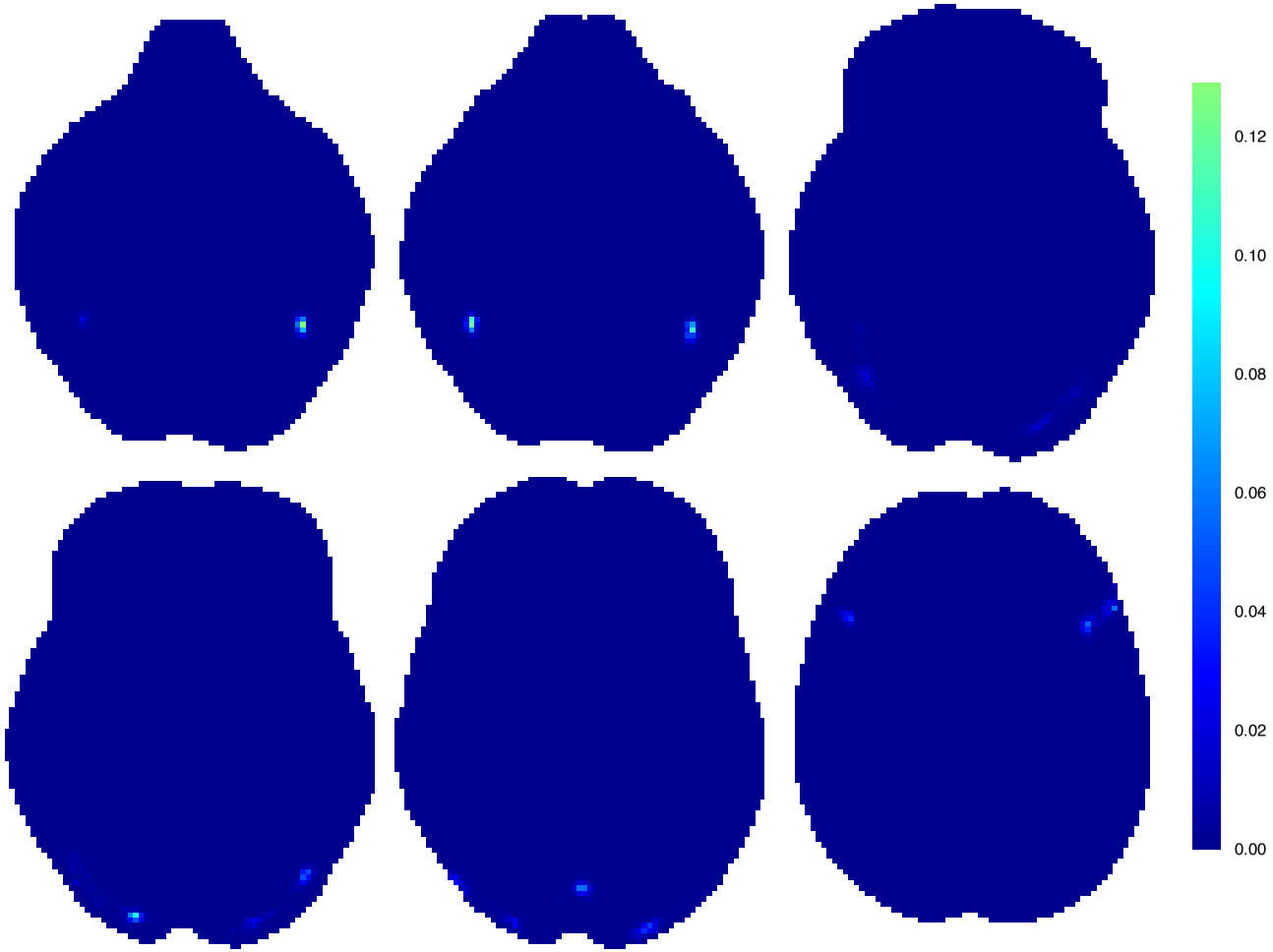



Figure B.6: Real data sensitivity analysis results. For each voxel, the figure shows the marginal posterior probabilities of observing an increase study center. In this run we use the same prior specifications defined in Section 5.5.2 except from $\eta$ which we set as 6 instead of 4.

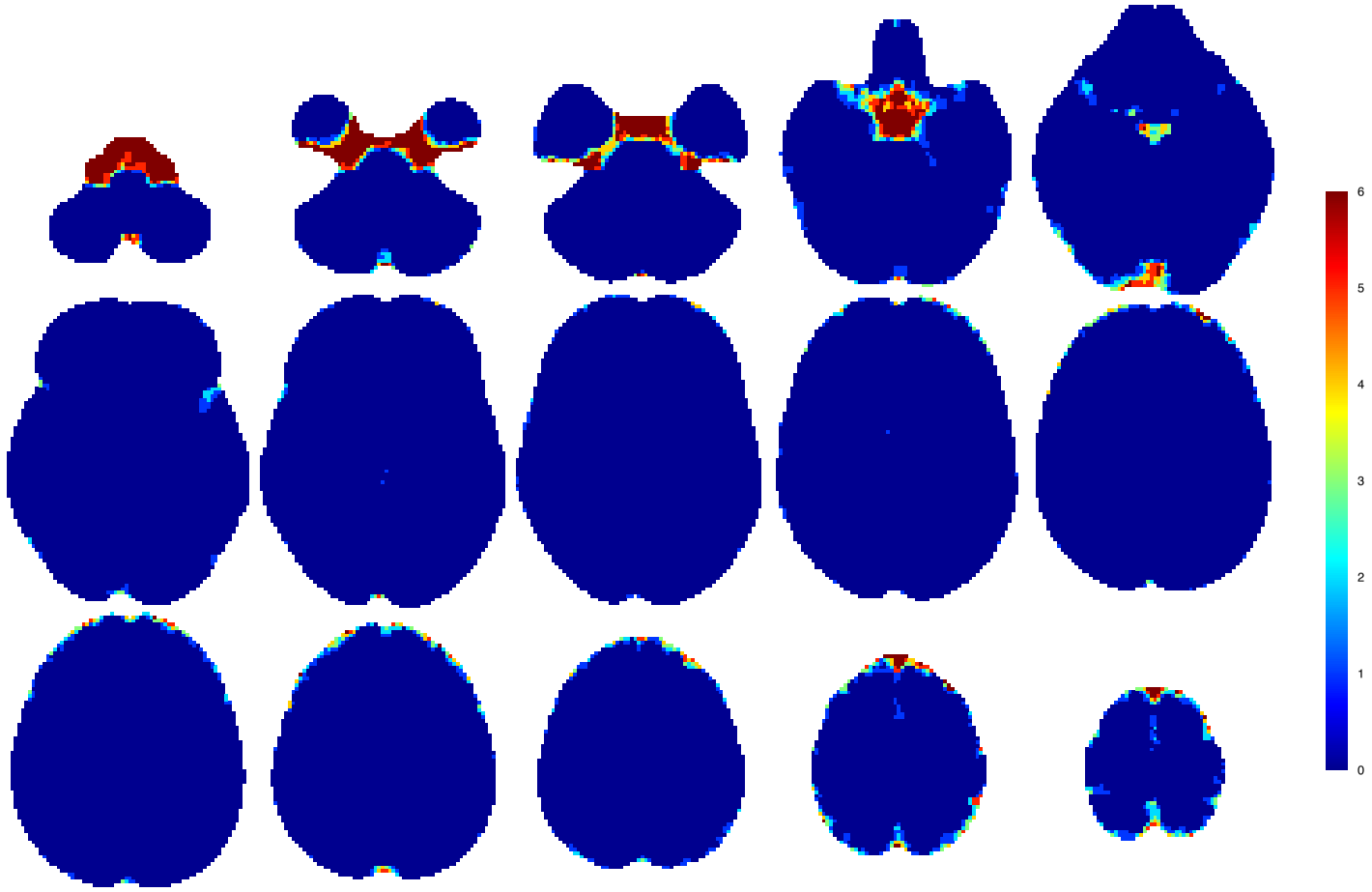



Figure B.7: Total number of participants that had missing data at a given voxel, for several axial slices. Top row (left to right): $z = -50$, $z = -42$, $z = -32$, $z = -24$ and $z = -16$. Middle row (left to right): $z = -8$, $z = +2$, $z = +10$, $z = +18$ and $z = +28$. Bottom row (left to right): $z = +36$, $z = +44$, $z = +52$, $z = +62$ and $z = +70$. Note that we found no study effects in regions where most of the participants had missing data.

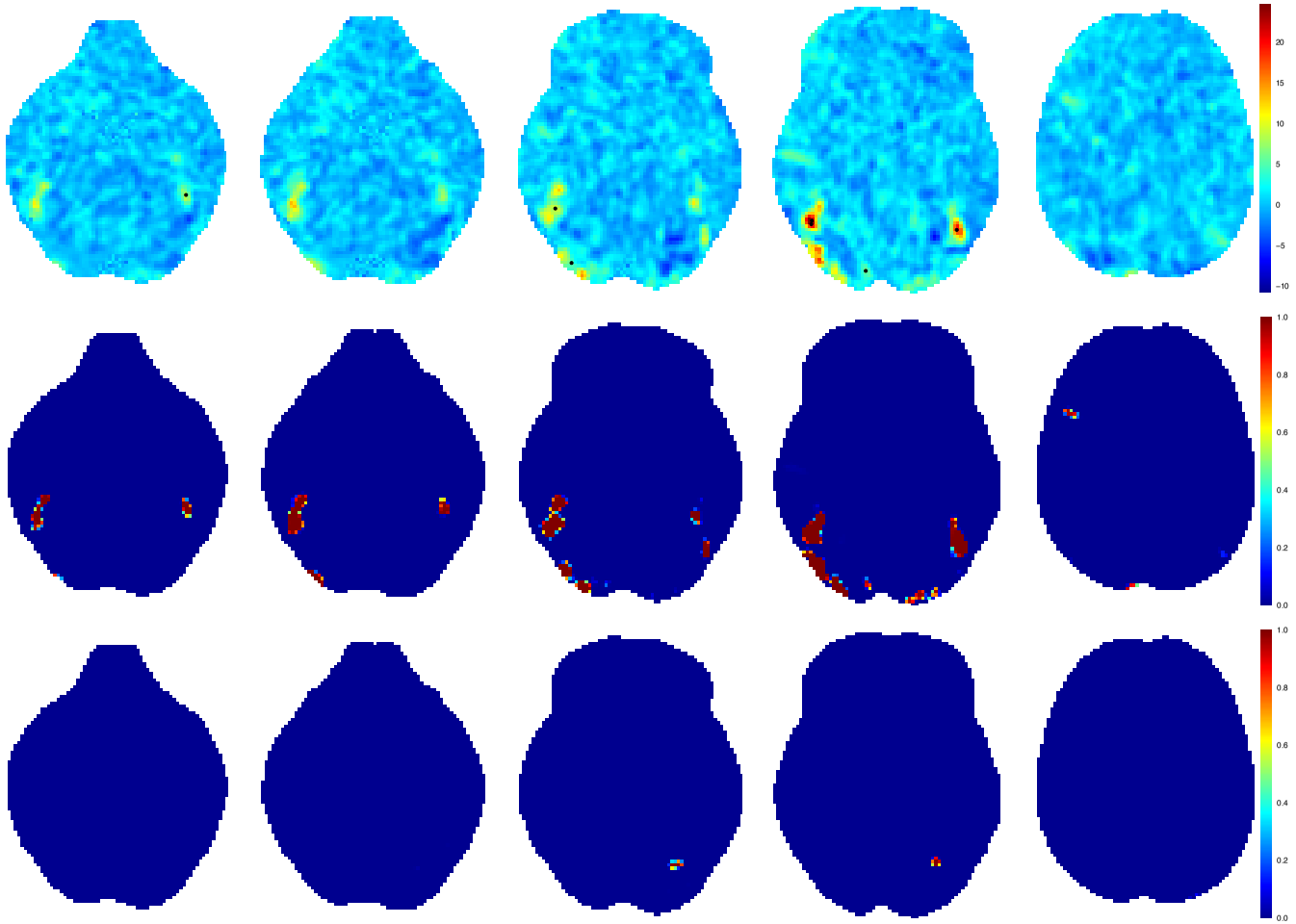



Figure B.8: Group fMRI study results for participant 2. The figure shows the observed data (top row), the voxel-wise probability of increase (middle row) and the voxel-wise probability of decrease (bottom row). Estimates are obtained based on 5,000 posterior draws. The black dots represent the individual center process $\mathbf{x}_1$ at iteration 100,000 of the MCMC. The columns correspond (from left to right) to axial slices $z = -20$, $z = -18$, $z = -14$, $z = -8$ and $z = +26$.

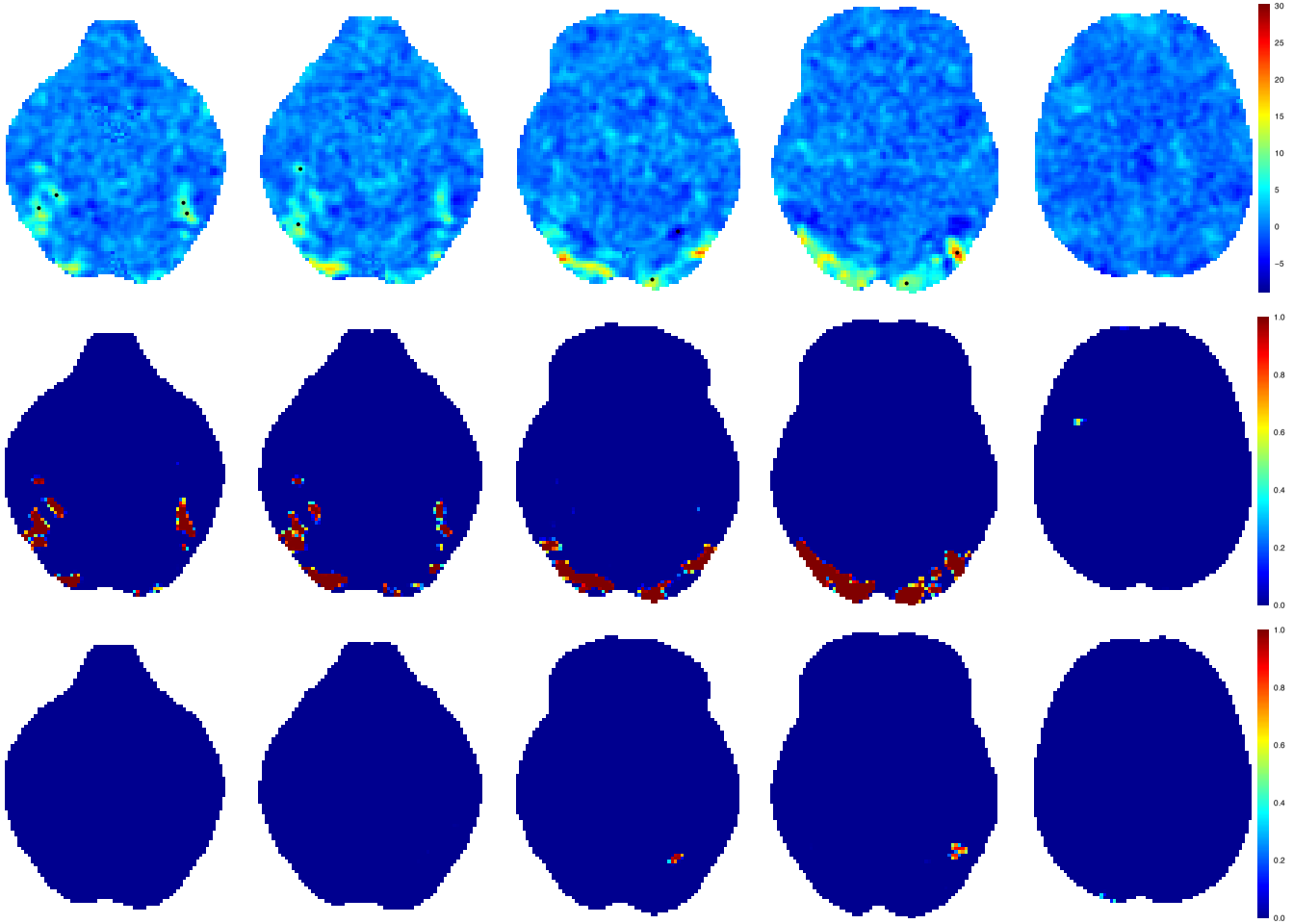



Figure B.9: Group fMRI study results for participant 3. The figure shows the observed data (top row), the voxel-wise probability of increase (middle row) and the voxel-wise probability of decrease (bottom row). Estimates are obtained based on 5,000 posterior draws. The black dots represent the individual center process $\mathbf{x}_1$ at iteration 100,000 of the MCMC. The columns correspond (from left to right) to axial slices $z = -20$, $z = -18$, $z = -14$, $z = -8$ and $z = +26$.

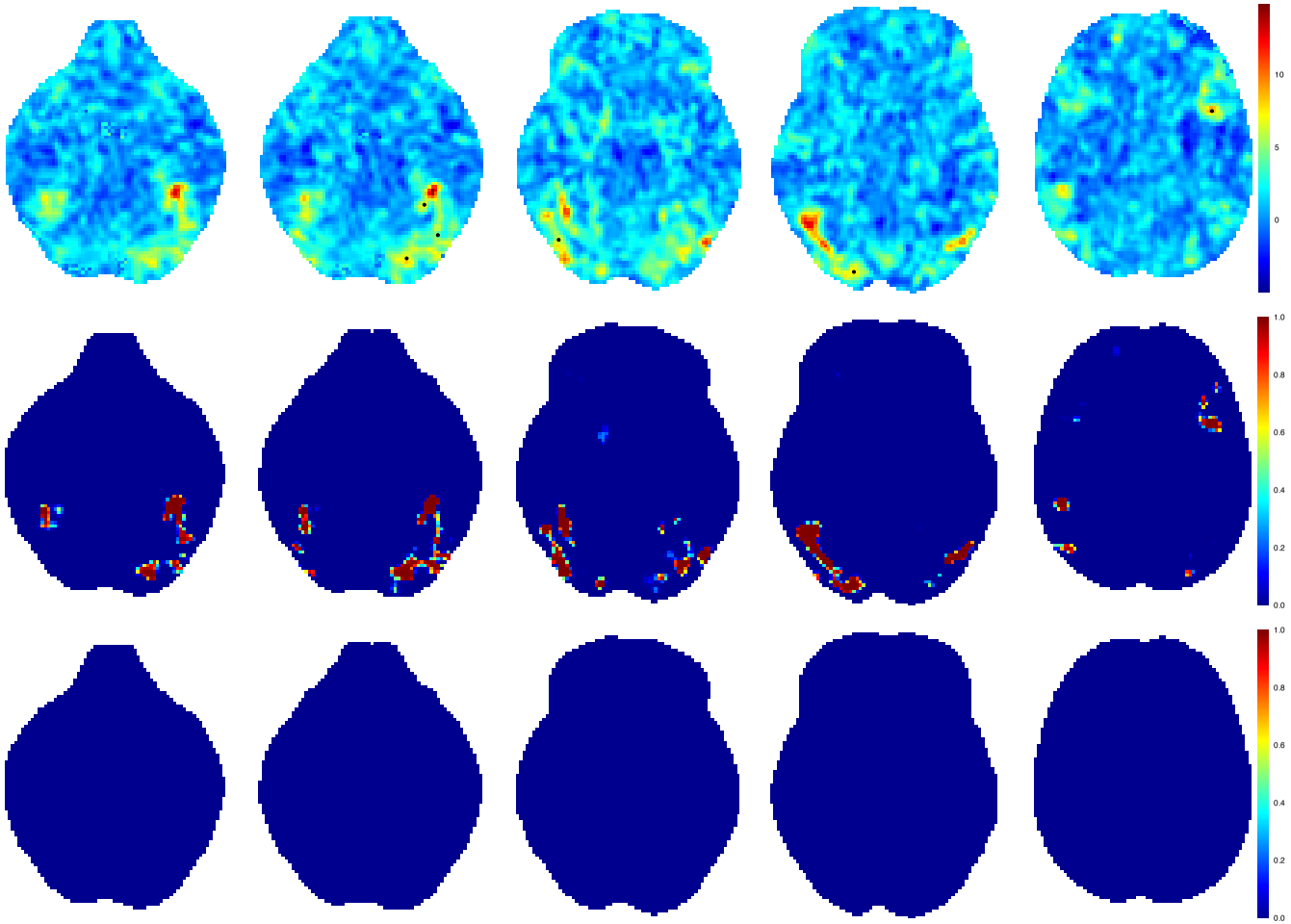



Figure B.10: Group fMRI study results for participant 4. The figure shows the observed data (top row), the voxel-wise probability of increase (middle row) and the voxel-wise probability of decrease (bottom row). Estimates are obtained based on 5,000 posterior draws. The black dots represent the individual center process $\mathbf{x}_1$ at iteration 100,000 of the MCMC. The columns correspond (from left to right) to axial slices $z = -20$, $z = -18$, $z = -14$, $z = -8$ and $z = +26$.

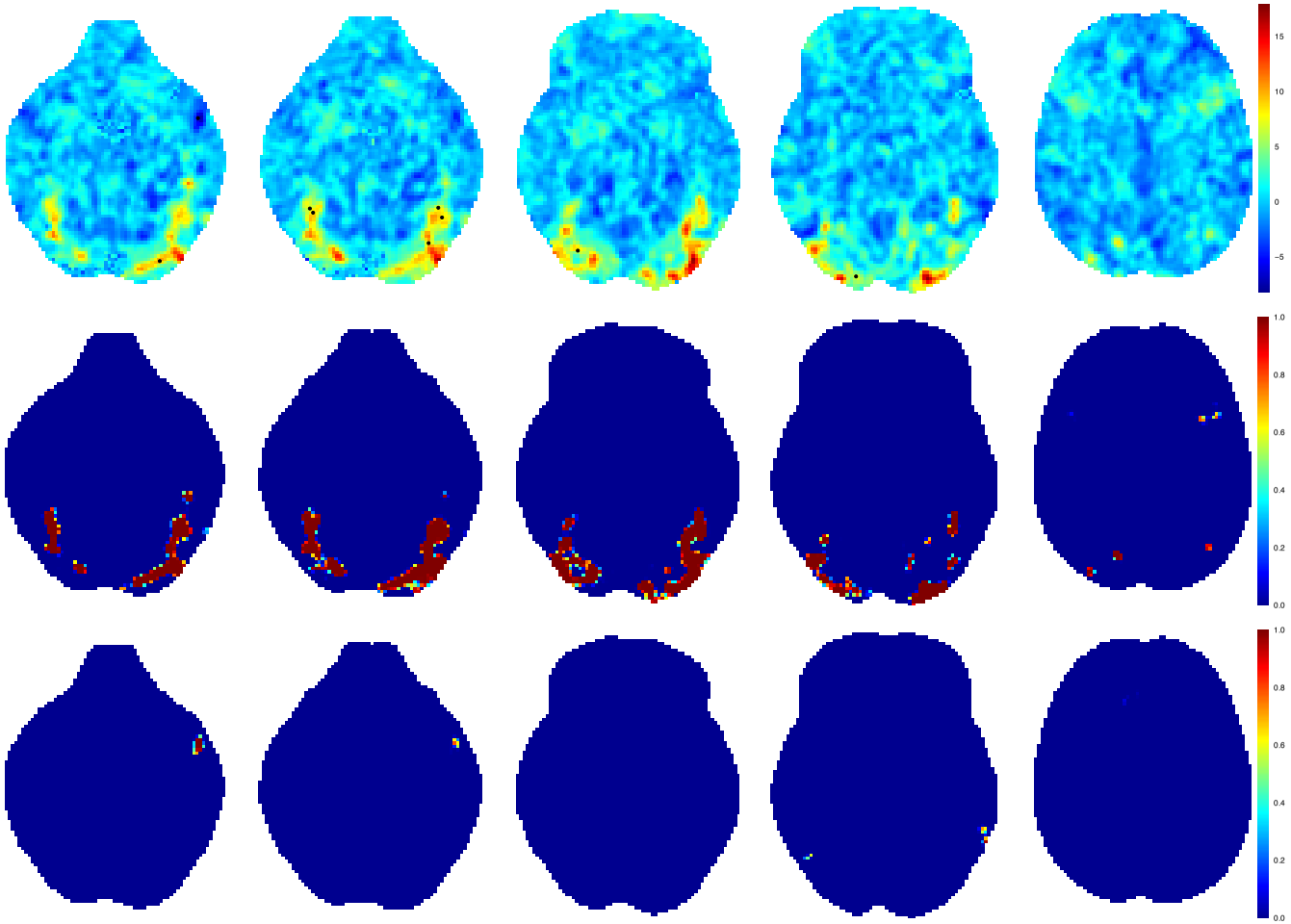



Figure B.11: Group fMRI study results for participant 5. The figure shows the observed data (top row), the voxel-wise probability of increase (middle row) and the voxel-wise probability of decrease (bottom row). Estimates are obtained based on 5,000 posterior draws. The black dots represent the individual center process $\mathbf{x}_1$ at iteration 100,000 of the MCMC. The columns correspond (from left to right) to axial slices $z = -20$, $z = -18$, $z = -14$, $z = -8$ and $z = +26$.

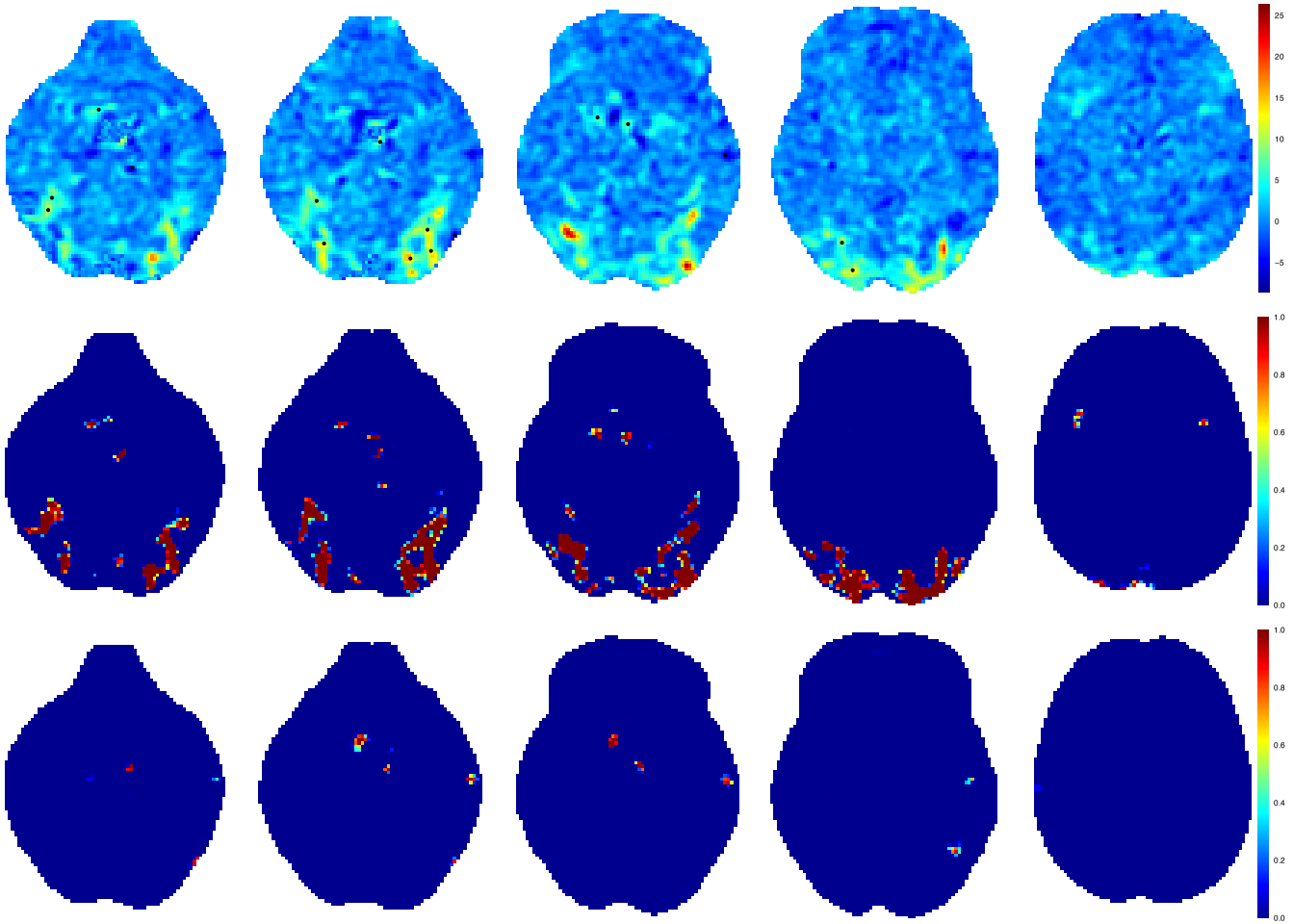



Figure B.12: Group fMRI study results for participant 6. The figure shows the observed data (top row), the voxel-wise probability of increase (middle row) and the voxel-wise probability of decrease (bottom row). Estimates are obtained based on 5,000 posterior draws. The black dots represent the individual center process $\mathbf{x}_1$ at iteration 100,000 of the MCMC. The columns correspond (from left to right) to axial slices $z = -20$, $z = -18$, $z = -14$, $z = -8$ and $z = +26$.

# Appendix C

# File drawer supplements

## C.1 Zero-truncated regression supplements

In this section we show results of zero-truncated regression models for subsamples B-E. Figure C.1 has the fitted zero-truncated Poisson and Negative Binomial models, with no covariates considered. In Figure C.2 we show estimated prevalence as a function of year of publication and study sample size. The effect of study context is demonstrated in Figure C.3. All quantities are calculated based on 100,000 MCMC samples, thinned at every 100 iterations. Because of the bad fit of the zero-truncated Poisson model we show regression results only for the zero-truncated Negative Binomial model fit.

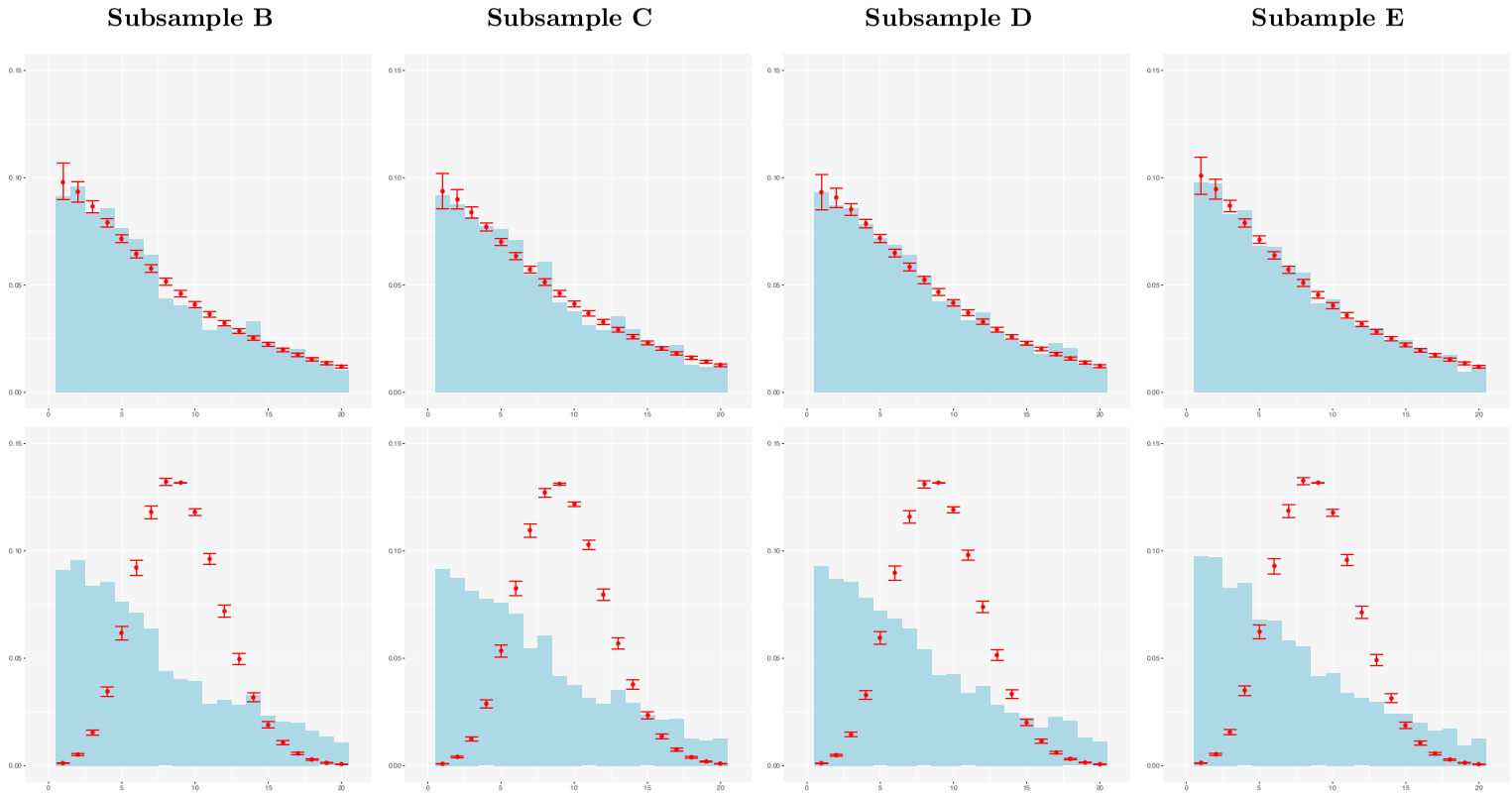



Figure C.1: Zero-truncated fit with the Negative Binomial (upper panel) and Poisson (lower panel) distributions for subsamples B-E. Points represent posterior medians obtained from 30,000 MCMC iterations, thinned every 30 iterations. Error bars represent the 95% credible intervals. Rows 1-4 correspond to subsamples B-E respectively.

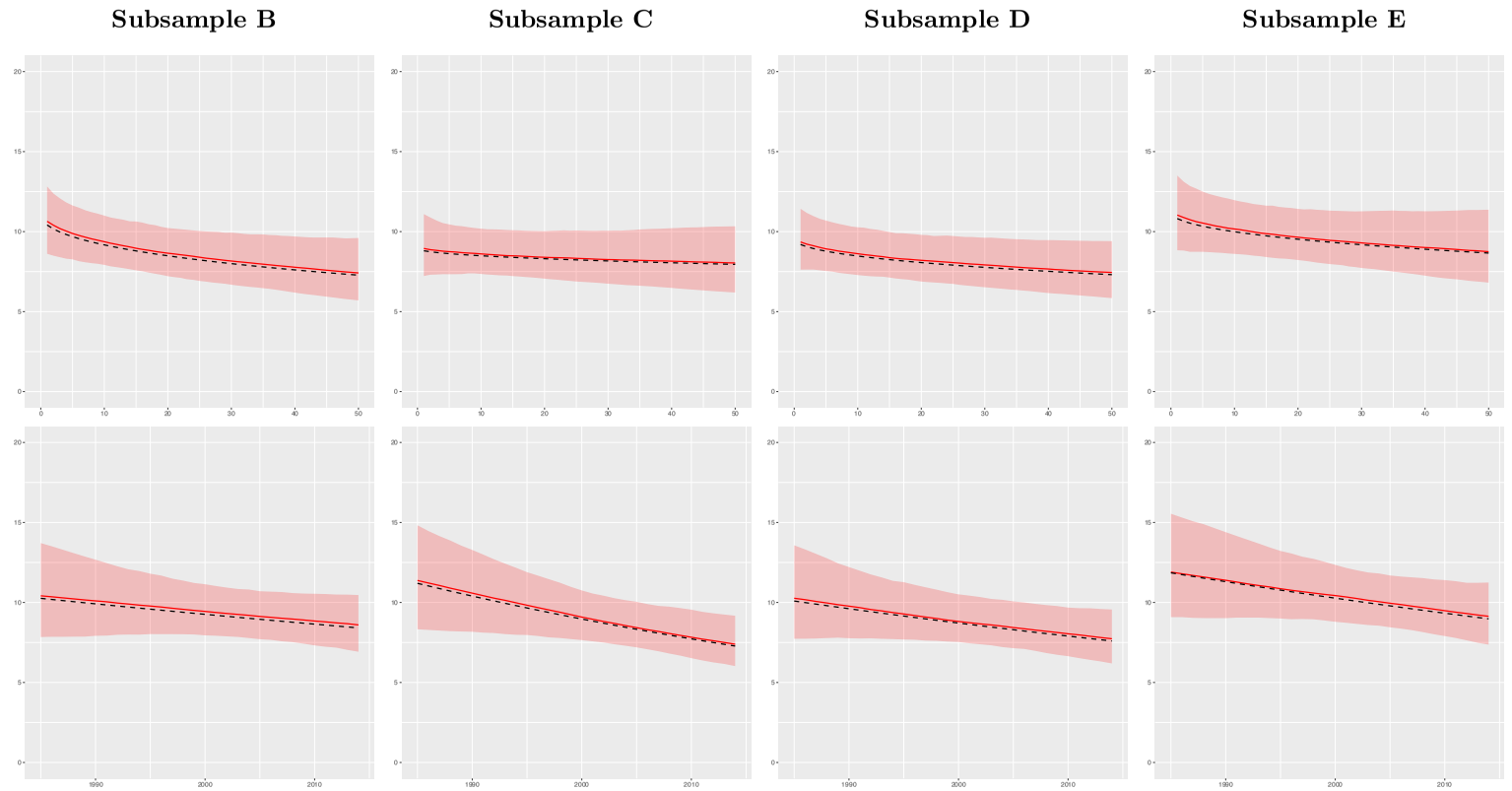



Figure C.2: Prevalence of file drawer studies as a function of study sample sample size (upper panel) and study year of publication (lower panel) for subsamples B-E. Red solid line is the posterior median whereas the shaded area indicates the 95% posterior credible intervals. The dashed line represents the estimate obtained with GAMLSS. All values are averages across the levels of study context.

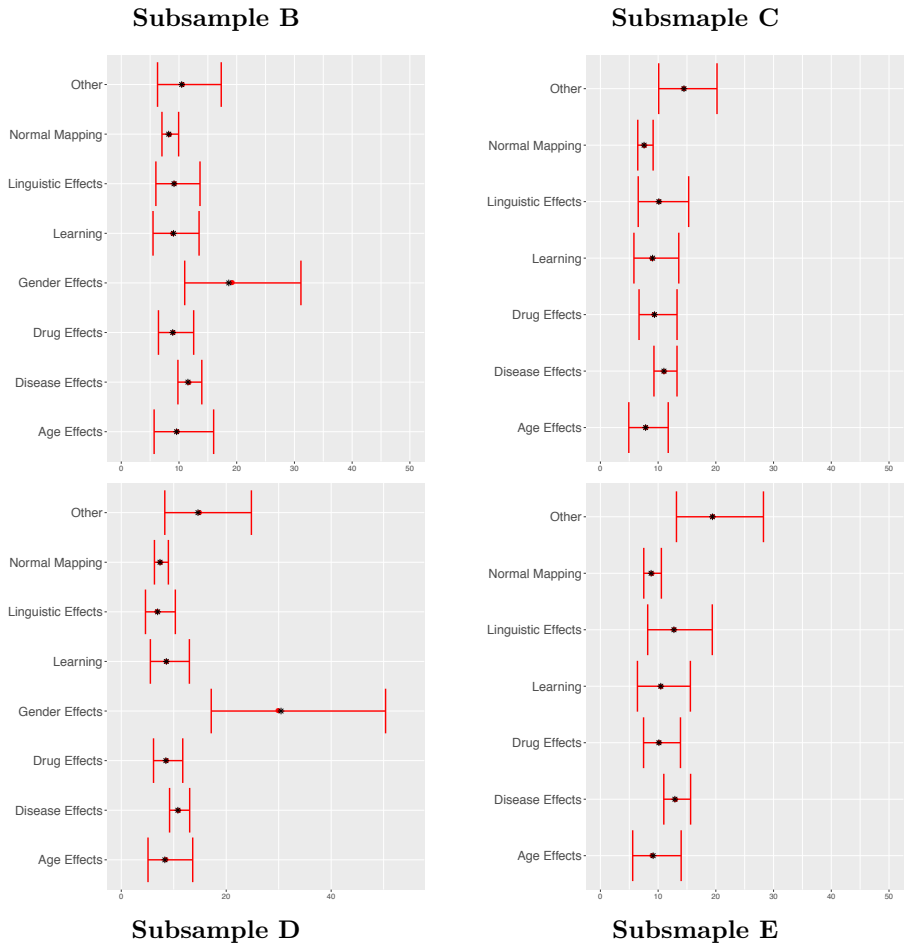



Figure C.3: Prevalence of file drawer studies as a function of study context for subsamples B-E. Red dots are the posterior medians whereas the error bars represent the 95% credible intervals. Black asterisks are the GAMLSS estimates. Publication year and sample size have been set to their median values.

## C.2 Emotion CBMA with missing studies

Figure C.4 shows median posterior intensities, obtained from the analysis of the emotion dataset presented in Section 4.5 after adding 86 zero foci studies. Analysis was done with the LGCP CBMA model presented in Chater 4; the estimates are based on a sample of 1,000 posterior draws from the HMC. Note that we use a different colorscale compared to Figure 4.6 which presents the results for the original dataset.

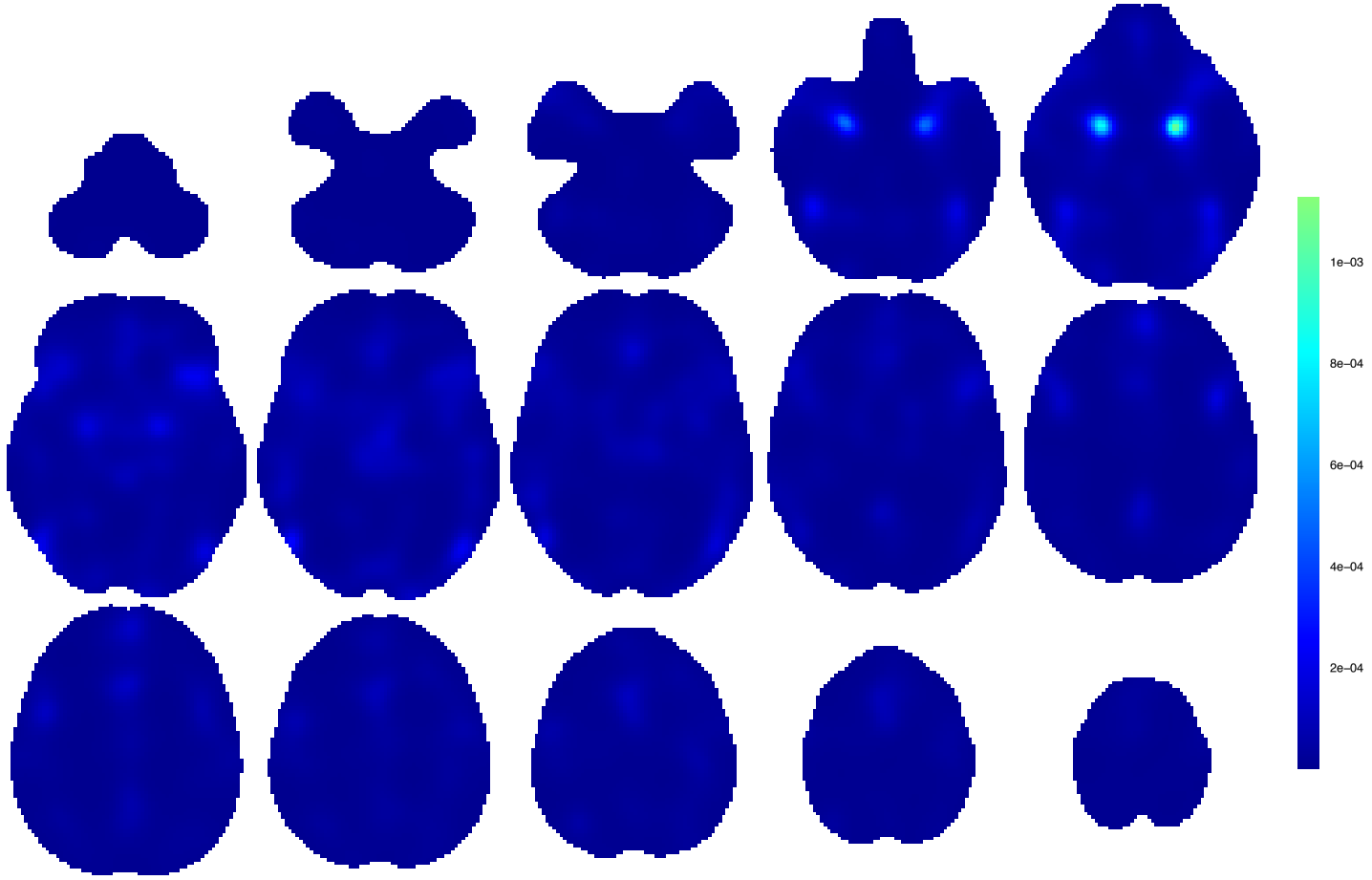



Figure C.4: Meta-analysis of the 855 emotion studies adding 86 zero-count studies. The figure presents the median posterior intensity at each voxel for several axial slices, as obtained from 1,000 independent draws from the posterior. Top row (left to right): $z = -50$, $z = -42$, $z = -32$, $z = -24$ and $z = -16$. Middle row (left to right): $z = -8$, $z = +2$, $z = +10$, $z = +18$ and $z = +28$. Bottom row (left to right): $z = +36$, $z = +44$, $z = +52$, $z = +62$ and $z = +70$.

## C.3 Zero-truncated Negative Binomial simulation study

In this section we perform a simulation study to assess the ability of the zero-truncated Negative Binomial model in estimating the probability of observing a zero-count study based on a sample from which these instances are removed. The zero-truncated Poisson model cannot be used for estimation of the file drawer quantity because it does not account for the overdispersion observed in the real data; hence, no simulation studies are presented.

Synthetic data are generated as follows. First, we choose the mean $\mu$ and the dispersion parameter $\phi$ of the Negative Binomial distribution. We then generate counts from the specified distribution so that the expected number of non-zero counts is $N$. Once the data have been generated we remove the zero-count instances and apply the MCMC to the remaining observations. Finally, we estimate the probability of observing a zero count study, $p_0$, based on our posterior samples.

Analysis of subsamples A-E yielded posterior distributions that were in the vicinity of 8.5 and 1.1 for $\mu$ and $\phi$, respectively; we therefore consider values 7-10 for the mean and values 0.8, 1, 1.2 and 1.5 for the dispersion. For each one the 16 combinations of $\mu$ and $\phi$, we sequentially set the expected number of observed studies to $N = 200,\ 500,\ 1,000$ and $2,000$ and for each case we generate 1,000 datasets. Overall, we have a total of 16,000 count samples to which we fit the zero-truncated Negative Binomial model. MCMC is run for 45,000 iterations in each distinct case; the first 15,000 of the draws are discarded as a burn-in and the remaining 30,000 are thinned every 30 iterations and subsequently used to estimate $p_0$.

The bias of the estimates $\hat{p}_0$ can be found in Table C.1 and the variance is shown in Table C.2. The results suggest that the bias of the estimate is negligible for all the parameter values considered in the study. The accuracy is generally good and improves as the sample size increases.

Table C.1: Results of the simulation study for the zero-truncated Negative Binomial model. Table shows bias of the estimates of $p_0$, the probability of observing a zero-count pattern based on a zero truncated sample. Bias is calculated based on 1,000 synthetic datasets. For each simulated dataset, the estimate of $p_0$ is obtained as the posterior median of $30,000$ MCMC iterations thinned every 30 iterations, after a burn-in of $15,000$ iterations. All values are multiplied by $10^3$.

| | | | **# observed studies** | | | |
|---|---|---|---|---|---|---|
| $\boldsymbol{\mu}$ | $\boldsymbol{\phi}$ | $\mathbf{p_0}$ | **200** | **500** | **1000** | **2000** |
| 7 | 0.8 | 0.162 | -0.70 | 1.32 | -0.12 | 0.25 |
| 8 | 0.8 | 0.147 | 0.20 | 1.33 | -0.06 | 0.01 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 9 | 0.8 | 0.135 | -0.01 | 0.28 | 0.06 | -0.08 |
| 10 | 0.8 | 0.125 | 0.75 | 0.17 | 0.18 | -0.07 |
| 7 | 1.0 | 0.125 | -0.23 | -0.31 | -0.90 | 0.75 |
| 8 | 1.0 | 0.111 | -0.15 | 0.17 | 0.02 | 0.39 |
| 9 | 1.0 | 0.100 | 0.36 | 0.28 | 0.32 | 0.06 |
| 10 | 1.0 | 0.091 | -0.44 | -0.44 | 0.25 | 0.26 |
| 7 | 1.2 | 0.100 | 0.80 | -0.14 | -0.10 | 0.05 |
| 8 | 1.2 | 0.087 | 0.22 | -0.18 | 0.07 | 0.05 |
| 9 | 1.2 | 0.077 | 0.30 | -0.08 | 0.07 | -0.14 |
| 10 | 1.2 | 0.069 | 0.24 | 0.77 | 0.25 | -0.01 |
| 7 | 1.5 | 0.074 | -0.43 | 0.58 | -0.13 | -0.03 |
| 8 | 1.5 | 0.063 | 0.19 | 0.75 | -0.02 | -0.11 |
| 9 | 1.5 | 0.054 | -0.07 | 0.50 | 0.15 | -0.14 |
| 10 | 1.5 | 0.047 | 0.19 | 0.42 | 0.18 | 0.07 |

Table C.2: Results of the simulation study for the zero-truncated Negative Binomial model. Table shows variance of the estimates of $p_0$, the probability of observing a zero-count pattern based on a zero truncated sample. Variance is calculated based on 1,000 synthetic datasets. For each simulated dataset, the estimate of $p_0$ is obtained as the posterior median of $30,000$ MCMC iterations thinned every 30 iterations, after a burn-in of $15,000$ iterations. All values are multiplied by $10^3$.

| | | | **# observed studies** | | | |
|---|---|---|---|---|---|---|
| $\boldsymbol{\mu}$ | $\boldsymbol{\phi}$ | $\mathbf{p_0}$ | **200** | **500** | **1000** | **2000** |
| 7 | 0.8 | 0.162 | 1.92 | 0.78 | 0.36 | 0.20 |
| 8 | 0.8 | 0.147 | 1.64 | 0.64 | 0.31 | 0.16 |
| 9 | 0.8 | 0.135 | 1.47 | 0.56 | 0.26 | 0.14 |
| 10 | 0.8 | 0.125 | 1.23 | 0.47 | 0.23 | 0.12 |
| 7 | 1.0 | 0.125 | 1.16 | 0.43 | 0.21 | 0.11 |
| 8 | 1.0 | 0.111 | 1.00 | 0.34 | 0.18 | 0.09 |
| 9 | 1.0 | 0.100 | 0.78 | 0.31 | 0.15 | 0.07 |
| 10 | 1.0 | 0.091 | 0.67 | 0.24 | 0.14 | 0.06 |
| 7 | 1.2 | 0.100 | 0.82 | 0.28 | 0.15 | 0.07 |
| 8 | 1.2 | 0.087 | 0.60 | 0.22 | 0.11 | 0.06 |
| 9 | 1.2 | 0.077 | 0.48 | 0.18 | 0.09 | 0.04 |
| 10 | 1.2 | 0.069 | 0.39 | 0.15 | 0.08 | 0.04 |
| 7 | 1.5 | 0.074 | 0.46 | 0.17 | 0.09 | 0.04 |
| 8 | 1.5 | 0.063 | 0.34 | 0.12 | 0.06 | 0.03 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 9 | 1.5 | 0.054 | 0.24 | 0.09 | 0.05 | 0.02 |
| 10 | 1.5 | 0.047 | 0.19 | 0.08 | 0.04 | 0.02 |

# Bibliography


Babu, G. J. and Feigelson, E. D. (1996). Spatial point processes in astronomy. *Journal of Statistical Planning and Inference*, **50**(3), 311–326. Spatial Statistics, Part I.

Bartels, A. and Zeki, S. (2004). The neural correlates of maternal and romantic love. *NeuroImage*, **21**(3), 1155–1166.

Begg, C. B. and Berlin, J. A. (1988). Publication bias: a problem in interpreting medical data. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, **151**(3), 419–463.

Begg, C. B. and Berlin, J. A. (1989). Publication bias and dissemination of clinical research. *Journal of the National Cancer Institute*, **81**(2), 107–115.

Begg, C. B. and Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, **50**(4), 1088–1101.

Benes, V., Bodlák, K., Møller, J., and Waagepetersen, R. P. (2002). Bayesian analysis of log Gaussian Cox processes for disease mapping. Technical report, Depertment of Mathematical Sciences, Aalborg University.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B: Methodological*, **57**(1), 289–300.

Bowman, F. D., Caffo, B., Bassett, S. S., and Kilts, C. (2008). A Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage*, **39**(1), 146–156.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., and Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews. Neuroscience*, **14**(5), 365–376.

Calhoun, V. D. and Pearlson, G. D. (2012). A selective review of simulated driving studies: combining naturalistic and hybrid paradigms, analysis approaches, and future directions. *NeuroImage*, **59**(1), 25–35.

Carp, J. (2012). The secret lives of experiments: Methods reporting in the fMRI literature. *NeuroImage*, **63**(1), 289–300.

Caspers, J., Zilles, K., Beierle, C., Rottschy, C., and Eickhoff, S. B. (2014). A novel meta-analytic approach: mining frequent co-activation patterns in neuroimaging databases. *NeuroImage*, **90**(0), 390–402.

Christensen, O. F. and Waagepetersen, R. P. (2002). Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics*, **58**(2), 280–286.

Christensen, O. F., Roberts, G. O., and Sköld, M. (2006). Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*, **15**(1), 1–17.

Cole, D. M., Smith, S. M., and Beckmann, C. F. (2010). Advances and pitfalls in the analysis and interpretation of resting-state fMRI data. *Frontiers in systems neuroscience*, **4**(8).

Copas, J. (1999). What works?: selectivity models and meta-analysis. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, **162**(1), 95–109.

Copas, J. (2013). A likelihood-based sensitivity analysis for publication bias in meta-analysis. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, **62**(1), 47–66.

Copas, J. and Jackson, D. (2004). A bound for publication bias based on the fraction of unpublished studies. *Biometrics*, **60**(1), 146–153.

Copas, J. and Shi, J. Q. (2000). Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*, **1**(3), 247–262.

Costafreda, S., Brammer, M., David, A., and Fu, C. (2008). Predictors of amygdala activation during the processing of emotional stimuli. *Brain Research Reviews*, **58**(1), 57–70.

Costafreda, S. G., David, A. S., and Brammer, M. J. (2009). A parametric approach to voxel-based meta-analysis. *NeuroImage*, **46**(1), 115–122.

Daley, D. and Vere-Jones, D. (2002). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Probability and Its Applications. Springer.

David, S. P., Ware, J. J., Chu, I. M., Loftus, P. D., Fusar-Poli, P., Radua, J., Munafò, M. R., and Ioannidis, J. P. A. (2013). Potential reporting bias in fMRI studies of the brain. *PLoS ONE*, **8**(7), e70104.

Davis, M. and Whalen, P. J. (2001). The amygdala: vigilance and emotion. *Molecular Psychiatry*, **6**(1), 13–34.

Delvecchio, G., Fossati, P., Boyer, P., Brambilla, P., Falkai, P., Gruber, O., Hietala, J., Lawrie, S. M., Martinot, J.-L., McIntosh, A. M., Meisenzahl, E., and Frangou, S. (2012). Common and distinct neural correlates of emotional processing in Bipolar Disorder and Major Depressive Disorder: a voxel-based meta-analysis of functional magnetic resonance imaging studies. *European Neuropsychopharmacology : the Journal of the European College of Neuropsychopharmacology*, **22**(2), 100–113.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, **31**(3), 968–980.

Dietrich, C. R. and Newsam, G. N. (1993). A fast and exact method for multidimensional Gaussian stochastic simulations. *Water Resources Research*, **29**(8), 2861–2869.

Diggle, P. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. CRC Press.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, **195**(2), 216–222.

Duval, S. and Tweedie, R. (2000a). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, **95**(449), 89–98.

Duval, S. and Tweedie, R. (2000b). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, **56**(2), 455–463.

Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A. W., Cronin, E., Decullier, E., Easterbrook, P. J., Von Elm, E., Gamble, C., Ghersi, D., Ioannidis, J. P. A., Simes, J., and Williamson, P. R. (2008). Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE*, **3**(8), e3081.

Dwan, K., Gamble, C., Williamson, P. R., Kirkham, J. J., and the Reporting Bias Group (2013). Systematic review of the empirical evidence of study publication bias and outcome reporting bias - An updated review. *PLoS ONE*, **8**(7), e66844.

Easterbrook, P. J., Berlin, J. A., Gopalan, R., and Matthews, D. R. (1991). Publication bias in clinical research. *Lancet*, **337**(8746), 867–876.

Eberly, L. E. and Casella, G. (1999). Bayesian estimation of the number of unseen studies in a meta-analysis. *Official Journal of Statistics*, **15**(4), 477–494.

Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, **315**(7109), 629–634.

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., and Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, **30**(9), 2907–2926.

Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., and Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, **59**(3), 2349–2361.

Etkin, A. and Wager, T. (2007). Functional neuroimaging of anxiety : a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *American Journal of Psychiatry*, **164**(10), 1476–1488.

Farah, M. J. (2014). Brain images, babies, and bathwater: critiquing critiques of functional neuroimaging. *The Hastings Center Report*, **44**(S2), S19–S30.

Flandin, G. and Penny, W. D. (2007). Bayesian fMRI data analysis with sparse spatial basis function priors. *NeuroImage*, **34**(3), 1108–1125.

Fox, P. T., Lancaster, J. L., Parsons, L. M., Xiong, J. H., and Zamarripa, F. (1997). Functional volumes modeling: theory and preliminary assessment. *Human Brain Mapping*, **5**(4), 306–311.

Fox, P. T., Parsons, L. M., and Lancaster, J. L. (1998). Beyond the single study: function/location metanalysis in cognitive neuroimaging. *Current Opinion in Neurobiology*, **8**(2), 178–187.

Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., and Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: theory. *NeuroImage*, **16**(2), 465–483.

Fusar-Poli, P. (2012). Voxel-wise meta-analysis of fMRI studies in patients at clinical high risk for psychosis. *Journal of Psychiatry and Neuroscience*, **37**(2), 106–112.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, **6**(4), 733–807.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis, Third Edition*. Taylor & Francis.

Genovese, C., Lazar, N., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, **15**(4), 870–878.

Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, **73**(2), 123–214.

Glimcher, P., Fehr, E., Camerer, C., and Poldrack, R. (2008). *Neuroeconomics: Decision Making and the Brain*. Elsevier Science.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4), 711–732.

Greenland, S. (1994). Invited commentary: a critical look at some popular meta-analytic methods. *American Journal of Epidemiology*, **140**(3), 290–296.

Haario, H., Saksman, E., and Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, **14**(3), 375–395.

Hariri, A. R., Tessitore, A., Mattay, V. S., Fera, F., and Weinberger, D. R. (2002). The amygdala response to emotional stimuli: a comparison of faces and scenes. *NeuroImage*, **17**(1), 317–323.

Harkness, R. D. and Isham, V. (1983). A bivariate spatial point pattern of ants' nests. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, **32**(3), 293–303.

Hartung, J., Knapp, G., and Sinha, B. K. (2008). *Statistical Meta-Analysis with Applications*. John Wiley & Sons, Hoboken.

Hartvig, N. and Jensen, J. (2000). Spatial mixture modeling of fMRI data. *Human Brain Mapping*, **11**(4), 233–248.

Hedges, L. V. (1989). Estimating the normal mean and variance under a publication selection model. In L. J. Gleser, M. D. Perlman, S. J. Press, and A. R. Sampson, editors, *Contributions to Probability and Statistics*, pages 447–458.

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science*, **7**(2), 246–255.

Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-analysis*. Academic Press.

Hill, A. C., Laird, A. R., and Robinson, J. L. (2014). Gender differences in working memory networks: A brainmap meta-analysis. *Biological Psychology*, **102**(0), 18–29.

Hoffman, M. and Gelman, A. (2014). The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, **15**(1), 1593–1623.

Huber, M. (2011). Spatial point processes. In S. Brooks, A. Gelman, G. L. Jones, and X. Meng, editors, *Handbook of Markov Chain Mote Carlo*, chapter 9, pages 227–252. Chapman & Hall/CRC.

Huettel, S. A., Song, A. W., and McCarthy, G. (2009). *Functional Magnetic Resonance Imaging Second Edition*. Sinauer Associates, Inc, Massachussets.

Illian, J., Penttinen, P., Stoyan, H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley.

Illian, J. B., Møller, J., and Waagepetersen, R. P. (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, **16**(3), 389–405.

Illian, J. B., Sørbye, S. H., and Rue, H. (2012a). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics*, **6**(4), 1499–1530.

Illian, J. B., Sørbye, S. H., Rue, H., and Hendrichsen, D. K. (2012b). Using INLA to fit a complex point process model with temporally varying effects – A case study. *Journal of Environmentl Statistics*, **3**(7), 1–25.

Iyengar, S. and Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science*, **3**(1), 133–135.

Jaakkola, T. and Jordan, M. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing*, **10**(1), 25–37.

Jin, Z., Zhou, X., and He, J. (2015). Statistical methods for dealing with publication bias in meta-analysis. *Statistics in Medicine*, **34**(2), 343–360.

Kang, J., Johnson, T. D., Nichols, T. E., and Wager, T. D. (2011). Meta analysis of functional neuroimaging data via Bayesian spatial point processes. *Journal of the American Statistical Association*, **106**(493), 124–134.

Kang, J., Nichols, T. E., Wager, T. D., and Johnson, T. D. (2014). A Bayesian hierarchical spatial point process model for multi-type neuroimaging meta-analysis. *The Annals of Applied Statistics*, **8**(3), 1561–1582.

Kanwisher, N. and Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **361**(1476), 2109–28.

Kim, S.-G. and Ogawa, S. (2012). Biophysical and physiological origins of blood oxygenation level-dependent fMRI signals. *Journal of Cerebral Blood Flow and Metabolism : Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, **32**(7), 1188–206.

Kirby, L. A. and Robinson, J. L. (2015). Affective mapping: An activation likelihood estimation (ALE) meta-analysis. *Brain and Cognition*. In press.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., and Wager, T. D. (2008). Functional grouping and cortical and subcortical interactions in emotion: a meta-analysis of neuroimaging studies. *NeuroImage*, **42**(2), 998–1031.

Konova, A. B., Moeller, S. J., and Goldstein, R. Z. (2013). Common and distinct neural targets of treatment: changing brain function in substance addiction. *Neuroscience and Biobehavioral Reviews*, **37**(10), 2806–2817.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., and Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, **89**(12), 5675–5679.

Laird, A. R., Lancaster, J. J., and Fox, P. T. (2005). Brainmap: the social evolution of a human brain mapping database. *Neuroinformatics*, **3**(1), 65–77.

Larose, D. T. and Dey, D. K. (1998). Modeling publication bias using weighted distributions in a Bayesian framework. *Computational Statistics and Data Analysis*, **26**(3), 279–302.

Lazar, N. A., Luna, B., Sweeney, J. A., and Eddy, W. F. (2002). Combining brains: a survey of methods for statistical pooling of information. *NeuroImage*, **16**(2), 538–550.

Liang, S., Carlin, B. P., and Gelfand, A. E. (2009). Analysis of minnesota colon and rectum cancer point patterns with spatial and nonspatial covariate information. *The Annals of Applied Statistics*, **3**(3), 943–962.

Light, R. J. and Pillemar, D. B. (1984). *Summing up: the science of reviewing research*. Harvard University Press.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., and Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences*, **35**(3), 121–143.

Lindquist, M. A. (2008). The statistical analysis of fMRI data. *Statistical Science*, **23**(4), 439–464.

Marshall, T. and Roberts, G. (2012). An adaptive approach to Langevin MCMC. *Statistics and Computing*, **22**(5), 1041–1057.

Møller, J. and Waagepetersen, R. P. (2003). An introduction to simulation-based inference for spatial point processes. In J. Møller, editor, *Spatial Statistics and Computational Methods*, chapter 4, pages 143–198. Springer-Verlag.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.

Møller, J. and Waagepetersen, R. P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, **34**(4), 643–684.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, **25**(3), 451–482.

Mumford, J. A. and Nichols, T. (2006). Modeling and inference of multisubject fMRI data. *Engineering in Medicine and Biology Magazine, IEEE*, **25**(2), 42–51.

Mumford, J. A. and Nichols, T. (2009). Simple group fMRI modeling and inference. *NeuroImage*, **47**(4), 1469–1475.

Murray, I., Adams, R. P., and MacKay, D. J. (2010). Elliptical slice sampling. *Journal of Machine Learning Research: Workshop and Conference Proceeding*, **9**(6), 541–548.

Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. L. Jones, and X. Meng, editors, *Handbook of Markov Chain Mote Carlo*, chapter 5, pages 113–162. Chapman & Hall/CRC.

Nichols, T. and Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, **12**(5), 419–446.

Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, **15**(1), 1–25.

Nielsen, F. Å. and Hansen, L. K. (2002). Modeling of activation data in the BrainMap$^{\text{TM}}$ database: detection of outliers. *Human Brain Mapping*, **15**(3), 146–156.

Normand, S. T. (1999). Meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine*, **18**(3), 321–359.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings*

*of the National Academy of Sciences of the United States of America*, **87**(24), 9868–9872.

Phelps, E. A. and LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, **48**(2), 175–187.

Poldrack, R. (2011). Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. *Neuron*, **72**(5), 692–697.

Poldrack, R. A., Mumford, J. A., and Nichols, T. (2011). *Handbook of Functional MRI Data Analysis*. Cambridge University Press.

Preston, C. J. (1977). Spatial birth-and-death processes. *Bulletin of the International Statistical Institute*, **46**(2), 371–391.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Radua, J. and Mataix-Cols, D. (2009). Voxel-wise meta-analysis of grey matter changes in obsessive-compulsive disorder. *The British Journal of Psychiatry : the Journal of Mental Science*, **195**(5), 393–402.

Radua, J. and Mataix-Cols, D. (2012). Meta-analytic methods for neuroimaging data explained. *Biology of Mood and Anxiety Disorders*, **2**(1), 6.

Radua, J., Mataix-Cols, D., Phillips, M. L., El-Hage, W., Kronhaus, D. M., Cardoner, N., and Surguladze, S. (2012). A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *European Psychiatry*, **27**(8), 605–611.

Radua, J., Rubia, K., Canales-Rodríguez, E. J., Pomarol-Clotet, E., Fusar-Poli, P., and Mataix-Cols, D. (2014). Anisotropic kernels for coordinate-based meta-analyses of neuroimaging studies. *Frontiers in Psychiatry*, **5**, 13.

Raemaekers, M., Vink, M., Zandbelt, B., van Wezel, R., Kahn, R., and Ramsey, N. (2007). Test-retest reliability of fMRI activation during prosaccades and antisaccades. *NeuroImage*, **36**(3), 532–542.

Raichle, M. E. (2003). Functional brain imaging and human brain function. *Journal of Neuroscience*, **23**(10), 3959–3962.

Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

Richlan, F., Kronbichler, M., and Wimmer, H. (2011). Meta-analyzing brain dysfunctions in dyslexic children and adults. *NeuroImage*, **56**(3), 1735–1742.

Rigby, R. A. and Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, **54**(3), 507–554.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, **86**(3), 638–641.

Rothstein, H. R., Sutton, A. J., and Borenstein, M. (2006). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Wiley.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., and Eickhoff, S. B. (2012). Modelling neural correlates of working memory: a coordinate-based meta-analysis. *NeuroImage*, **60**(1), 830–846.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D., and Nichols, T. E. (2009). Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *NeuroImage*, **45**(3), 810–823.

Shermer, M. (2008). Why you should be skeptical of brain scans. *Scientific American Mind*, **19**(5), 66–71.

Simpson, D., Illian, J., Lindgren, F., Sørbye, S., and Rue, H. (2016). Going off grid: computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, **103**(1), 49–70.

Song, F., Eastwood, A. J., Gilbody, S., Duley, L., and Sutton, A. J. (2000). Publication and related biases. *Health Technology Assessment*, **4**(10), 1–191.

Spiegelhalter, D., Abrams, K., and Myles, J. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Statistics in Practice. Wiley.

Stasinopoulos, D. M. and Rigby, R. A. (2007). Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, **23**(7), 1–46.

Sterling, T. D., Rosenbaum, W. L., and Weinkam, J. J. (1995). Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, **49**(1), 108–112.

Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., and Jones, D. R. (2000a). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, **320**(7249), 1574–1577.

Sutton, A. J., Song, F., Gilbody, S. M., and Abrams, K. R. (2000b). Modelling publication bias in meta-analysis: a review. *Statistical Methods in Medical Research*, **9**(5), 421–445.

Talairach, J. and Tournoux, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme, Stuttgart.

Taylor, B. M. and Diggle, P. J. (2014). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation*, **84**(10), 2266–2284.

Thirion, B., Pinel, P., Meriaux, S., Roche, A., Dehaene, S., and Poline, J.-B. (2007). Analysis of a large fMRI cohort: statistical and methodological issues for group analyses. *NeuroImage*, **35**(1), 105–120.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., and Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage*, **16**(3, Part A), 765–780.

Turkeltaub, P. E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., and Fox, P. (2012). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping*, **33**(1), 1–13.

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Penna, S. D., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S., Prior, F., Schlaggar, B., Smith, S., Snyder, A., Xu, J., and Yacoub, E. (2012). The Human Connectome Project: a data acquisition perspective. *NeuroImage*, **62**(4), 2222–2231.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., and Ugurbil, K. (2013). The WU-Minn Human Connectome Project: an overview. *NeuroImage*, **80**, 62–79.

van Houwelingen, H. C., Arends, L. R., and Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*, **21**(4), 589–624.

van Lieshout, M. N. M. and Baddeley, A. J. (2002). Extrapolating and interpolating spatial patterns. In A. Lawson and D. Denison, editors, *Spatial Cluster Modelling*, pages 61–86. Chapman & Hall/CRC.

Vul, E., Harris, C., Winkielman, P., and Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, **4**(3), 274–290.

Waagepetersen, R. and Schweder, T. (2006). Likelihood-based inference for clustered line transect data. *Journal of Agricultural, Biological, and Environmental Statistics*, **11**(3), 264–279.

Waagepetersen, R. P. (2004). Convergence of posteriors for discretized log Gaussian Cox processes. *Statistics and Probability Letters*, **66**(3), 229–235.

Wager, T. D., Phan, K., Liberzon, I., and Taylor, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *NeuroImage*, **19**(3), 513–531.

Wager, T. D., Jonides, J., and Reading, S. (2004). Neuroimaging studies of shifting attention: a meta-analysis. *NeuroImage*, **22**(4), 1679–1693.

Wager, T. D., Lindquist, M., and Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: current and future directions. *Social Cognitive and Affective Neuroscience*, **2**(2), 150–158.

Wager, T. D., Lindquist, M. a., Nichols, T. E., Kober, H., and Van Snellenberg, J. X. (2009). Evaluating the consistency and specificity of neuroimaging data using meta-analysis. *NeuroImage*, **45**(Supplement 1), S210—-S221.

Wolpert, R. L. and Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika*, **85**(2), 251–267.

Wood, A. T. A. and Chan, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*, **3**(4), 409–432.

Xu, L., Johnson, T. D., Nichols, T., and Nee, D. E. (2009). Modeling inter-subject variability in fMRI activation location: A Bayesian hierarchical spatial model. *Biometrics*, **65**(4), 1041–1051.

Xue, W., Kang, J., Bowman, F. D., Wager, T. D., and Guo, J. (2014). Identifying functional co-activation patterns in neuroimaging studies via poisson graphical models. *Biometrics*, **70**(4), 812–822.

Yarkoni, T., Poldrack, R. A., Van Essen, D. C., and Wager, T. D. (2010). Cognitive neuroscience 2.0: building a cumulative science of human brain function. *Trends in Cognitive Sciences*, **14**(11), 489–496.

Yue, Y. and Speckman, P. L. (2010). Nonstationary spatial Gaussian Markov random fields. *Journal of Computational and Graphical Statistics*, **19**(1), 96–116.

Yue, Y. R., Lindquist, M. A., and Loh, J. M. (2012). Meta-analysis of functional neuroimaging data using Bayesian nonparametric binary regression. *The Annals of Applied Statistics*, **6**(2), 697–718.

Zurawicki, L. (2010). *Neuromarketing: Exploring the Brain of the Consumer*. Springer Berlin Heidelberg.